DOCUMENT RESUME

ED 137 418 TM 006 232

AUTHOR Cannel, Charles F.; And Others
TITLE A Summary of Studies of Interviewing Methodology. Vital and Health Statistics: Series 2, Data Evaluation and Methods Research; No. 69. DHEW Publication No. (HRA) 77-1343.
INSTITUTION Michigan Univ., Ann Arbor. Survey Research Center.; National Center for Health Statistics (DHEW), Rockville, Md.
PUB DATE Mar 77
NOTE 90p.
AVAILABLE FROM Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 10401 (Stock no. 017-022-00533-6, $1.45)

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS Bias; *Data Collection; Feedback; *Interviews; Memory; Problems; *Public Health; Questioning Techniques; Questionnaires; Reinforcement; *Research Reviews (Publications); Response Mode; Sampling; Supervision; *Surveys; Training; Validity; Verbal Communication

ABSTRACT

In several studies of experimental interviewing techniques and their effect on reporting behavior described in this publication, an attempt is made to identify the elements of the interview process that are potential sources for improving data collection. Methodological studies designed to test the effectiveness of certain questionnaire designs and interviewing techniques used in the collection of data on health events in household interviews are presented. The role of behaviors, attitudes, perceptions, and information levels of both the respondent and the interviewer is investigated. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

VITAL and HEALTH STATISTICS
DATA EVALUATION AND METHODS RESEARCH
Series 2-No. 69

ED137418

NATIONAL CENTER FOR HEALTH STATISTICS
NCHS

A Summary of Studies of Interviewing Methodology

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
Public Health Service
Health Resources Administration

2

Library of Congress Cataloging in Publication Data

Cannell, Charles F

A summary of research studies of interviewing methodology, 1959-1970.

(Vital and health statistics: Series 2, Data evaluation and methods research; no. 69) (DHEW publication; no. (HRA) 77-1343)

Supt. of Docs. no.: HE 20.6209:2/69

Bibliography: p.

1. Health surveys. 2. Interviewing. 3. Medical history taking. I. Marquis, Kent H., joint author. II. Laurent, André, joint author. III. Title. IV. Series: United States. National Center for Health Statistics. Vital and health statistics: Series 2, Data evaluation and methods research; no. 69. V. Series: United States. Dept. of Health, Education, and Welfare. DHEW publication; no. (HRA) 77-1343. [DNLM: 1. Medical history taking. W2 AN148vb no. 69]

RA 409.U45 no. 69 312'.07'23s [312'.07'23]

ISBN 0-8406-0062-3 75-619406

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $1.45

Stock No. 017-022-00623-6

DATA EVALUATION AND METHODS RESEARCH

Series 2
Number 69

A Summary of Studies
of Interviewing Methodology

A summary of methodological studies designed to test the effectiveness of certain questionnaire designs and interviewing techniques used in the collection of data on health events in household interviews and to investigate the role of behaviors, attitudes, perceptions, and information levels of both the respondent and the interviewer.

DHEW Publication No. (HRA) 77-1343

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Public Health Service

Health Resources Administration
National Center for Health Statistics
Rockville, Md. March 1977

NATIONAL CENTER FOR HEALTH STATISTICS

DOROTHY P. RICE, *Director*

ROBERT A. ISRAEL, *Deputy Director*
JACOB J. FELDMAN, Ph.D., *Associate Director for Analysis*
GAIL F. FISHER, *Associate Director for the Cooperative Health Statistics System*
ELIJAH L. WHITE, *Associate Director for Data Systems*
ANDERS S. LUNDE, Ph.D., *Associate Director for International Statistics*
ROBERT C. HUBER, *Associate Director for Management*
MONROE G. SIRKEN, Ph.D., *Associate Director for Mathematical Statistics*
PETER L. HURLEY, *Associate Director for Operations*
JAMES M. ROBEY, Ph.D., *Associate Director for Program Development*
PAUL E. LEAVERTON, Ph.D., *Associate Director for Statistical Research*
ALICE HAYWOOD, *Information Officer*

DIVISION OF HEALTH INTERVIEW STATISTICS

ROBERT R. FUCHSBERG, *Director*
PETER RIES, Ph.D., *Chief, Illness and Disability Statistics Branch*
ROBERT A. WRIGHT, *Acting Chief, Utilization and Expenditure Statistics Branch*
CLINTON E. BURNHAM, *Chief, Survey Planning and Development Branch*

COOPERATION OF THE U.S. BUREAU OF THE CENSUS

In accordance with specifications established by the National Health Survey, the Bureau of the Census, under a contractual agreement, participated in the design and selection of the sample, and carried out the first stage of the field interviewing and certain parts of the statistical processing.

Vital and Health Statistics - Series 2 - No. 69

DHEW Publication No. (HRA) 77-1343
Library of Congress Catalog Card Number 75-619406

PREFACE

For more than a decade the Survey Research Center of the University of Michigan and the Division of Health Interview Statistics of the National Center for Health Statistics (NCHS) have had a continuous contractual arrangement for the investigation of response problems in reporting health information in sampling surveys.

The contract program, which started in the late 1950's shortly after the initiation of the Health Interview Survey, began with a series of validity studies in which samples drawn from medical records were compared with data collected by interview. These studies were designed to identify patterns of response bias as a basis for developing procedures to improve reporting. These investigations of levels of underreporting and characteristics of response patterns were evaluated in terms of respondent status, the attitudes and behavior of the interviewer and the respondent, and nature of the events being reported. These studies are discussed in the sections, "Behavior in Interviews," and "Interviewer Performance Difference" of this publication.

The more recent studies, which have developed out of findings of the preceding research, involved experimental procedures designed to improve reporting. Investigated were such procedures as the use of verbal reinforcement of the respondent, probing as a method of improving memory and information retrieval, and varying the length of questions in an attempt to increase respondent participation in the interview. These studies are described in the sections "The Use of Verbal Reinforcement in Interviews and Its Data Accuracy," "Memory and Information Retrieval in the Interview," and "Question Length and Reporting Behavior in the Interview" of this publication.

All but one of the NCHS studies summarized in this report have appeared as complete research reports in series 2 of *Vital and Health Statistics*. The study by Cannell and Fowler (1963)[1] on validity of reporting visits to physicians was not published in the series. The second study of interviewer-respondent interactions by Marquis and Cannell[2] was done in 1969 under a contract with the Department of Labor, Manpower Administration. All others were contracted for by NCHS. In addition, two NCHS studies not conducted by the Survey Research Center are frequently referred to here because they had as their subject some of the same problems of reporting: one in 1967 by W. G. Madow, the Stanford Research Institute report published as Series 2-Number 23[3]; and the other in 1961 by E. Balamuth, et al., the Health Insurance Plan study, most recently published as Series 2-Number 7[4].

Because these studies have had considerable interest for methodologists and for survey researchers more generally, it was

thought useful to review them in a single volume so that the sequence of the major lines of inquiry could be followed. This report does not include a review of literature nor does it attempt to integrate underlying theories. It does present the findings in such a way as to make apparent their consistencies or inconsistencies, and does discuss some underlying hypotheses. This compilation also allows more emphasis to be placed on interpretation and explanation than was possible in the individual presentations.

In the concluding sections of this report the findings of the several studies are synthesized, a model of reporting is developed, and a description is offered of how the research performed at the Survey Research Center (SRC) has been applied to collection procedures used in the Health Interview Survey (HIS) to improve the quality of the collected data.

Since these studies were completed, much additional methodological work has been conducted by the Survey Research Center focusing on experimental procedures for improving the validity of reporting. This newer research at times confirms findings in this report, provides further support for these hypotheses and, at times, runs counter to some of the conclusions. Some of these findings can be found in a forthcoming report, "Experiment in Interviewing Techniques," summarizing research conducted by SRC for the National Center for Health Services Research.

The contractual relationship between the SRC and the HIS does not consist solely of a financial arrangement; much of the research is the cooperative work of the two organizations. The Bureau of the Census has also been an active participant in several of the studies, both in the planning and data collection phases.

Charles F. Cannell
Program Director,
Survey Research Center,
University of Michigan

ACKNOWLEDGMENTS

In a cooperative and integrated research program it is difficult to acknowledge the contribution of all participants. This is especially true of this project because the studies have extended more than a decade and the research staff has changed during that time. In addition to the authors, others who have had major responsibility in one or more of the studies include Mr. Thomas Bakker, Professor Gordon Fisher, Floyd Fowler, Ph.D., and Thomas deKoning, Ph.D. Special assistance in the preparation of the manuscripts was provided by Linda Winter and Marion Wirick.

SYMBOLS USED IN TABLES

Data not available	. . .
Category not applicable	. . .
Quantity zero	-
Quantity more than 0 but less than 0.05	0.0
Figure does not meet standards of reliability or precision	*

CONTENTS

	Page
Preface	iii
Acknowledgments	v
Introduction	1
Understanding the Interview Process	1
Influence of the Interviewer	1
Response Error	2
Problems of Recall and Information Retrieval	2
Bias Introduced Through Interviewer Feedback	3
Research Needs	4
Studies of Underreporting of Health Events in the Household Interview	4
Underreporting and Characteristics of Health Events	5
Effect of Elapsed Time on Reporting	5
Effect of Impact of the Event Upon Reporting	8
Effect of Social and Personal Threat Upon Reporting	9
Summary	11
Underreporting and Characteristics of Respondents	11
Age of Respondent	11
Sex of Respondent	13
Education of Respondent	13
Family Income of Respondent	14
Color of Respondent	15
Reporting for Self Versus Reporting for Other Family Member	15
Conclusions	16
Behavior in Interviews	17
Health Interview Survey Observation Study	17
Health Interview Survey Data	17
Observation	18
Interviewer Ratings of the Respondent	19
Reinterview With the Respondent	19
Interview With the Health Interviewer	19
Results	19
Urban Employment Survey Behavior Interaction Study	22
New Coding Scheme	22
Main Findings	22
Interviewer Performance Difference: Some Implications for Field Supervision and Training	30
Conclusions and Discussion	36

Page

The Use of Verbal Reinforcement in Interviews and Its Data Accuracy 37
Research on Interviewer Reinforcement 38
Operant Conditioning Studies With Verbal Reinforcement 38
Effects of Feedback on Amount Reported 38
Use of Feedback To Increase Accuracy 39
Use of Feedback in a Nonexperimental Interview 42
Discussion 42
Three Kinds of Interviewer Verbal Reinforcement Effects 43
Cognitive Effects of Reinforcement 43
Conditioning Effects of Reinforcement 46
Motivational Effects of Reinforcement 48
Experimental Study 50

Memory and Information Retrieval in the Interview 52
The Inadequate Search Hypothesis 52
An Integrative Hypothesis: Cognitive Inadequacy of Stimuli Questions 53
Design of an Experimental Interviewing Approach 53
The Extensive Questionnaire 55
Control Questionnaire 55
Field Experiment 56
Dependent Variables 56
Hypothesis 56
Results and Discussion 57
Conclusions 59

Question Length and Reporting Behavior in the Interview: Preliminary Investigations 60
Empirical Findings on Behavior Matching in the Interview 61
Hypotheses About the Effects of Question Length on Reporting Behavior 61
Experiment I: Effects of Question Length on Answer Duration and Reporting Frequency 62
Questionnaire Procedures 62
Field Procedures 63
Dependent Variables 64
Results and Discussion 64
Experiment II: Effects of Question Length on Validity of Report 66
Results and Discussion 67
Conclusions 70

References 72

Appendix. Application of Survey Research Center Findings to Health Interview Survey Procedures 75
Recall of Health Events 75
Effective Probing for Health Events 76
Interviewer-Respondent Communication 77
Other Considerations 77

A SUMMARY OF STUDIES OF INTERVIEWING METHODOLOGY

Charles F. Cannell, Kent H. Marquis, and André Laurent,
Survey Research Center, Institute for Social Research The University of Michigan

INTRODUCTION

Survey interviewing as a technique of data collection has developed from early attempts to collect simple demographic information to the current more sophisticated inquiries concerning attitudes, motives, and a wide variety of factual information. Despite the increasingly complex demands on the survey interview, methodologies for question construction and interviewer behavior have not changed a great deal.

Much research on interview method has been directed to the general problems of underreporting, to inaccuracies in interview data due to interviewer bias or response error, and to the problems of recall and information retrieval. The inadequacies of the interview method have been well documented and the need for improved techniques in data collection is readily apparent. However, little has been done toward perfecting the interview procedure as a method of data collection.

One reason interviewing techniques have advanced slowly may be that interviewing has no comprehensive theory to draw upon for cause-and-effect relationships. Ideas about effective questioning must be drawn from fragments of psychological theory or, more often, from folklore, experience, and common sense. Before major advances can be made, it is necessary to learn more about what happens in the interview situation and to develop some theories about the cause-and-effect sequences that occur. In several studies of experimental interviewing techniques and their effect on reporting behavior described in this publication, an attempt has been made to identify the elements of the interview process that are potential sources for improving data collection.

UNDERSTANDING THE INTERVIEW PROCESS

Influence of the Interviewer

Early attempts to investigate inaccurate reporting in interview surveys were focused primarily on the interviewer. Results of these early studies suggested that the interviewer's attitudes, expectations, background, and physical characteristics introduced important sources of bias into the household interview.

In a 1929 pioneer study of interviewer effect reported by Stuart Rice,[5] it was found that interviewers who were prohibitionists were likely to ascribe the sad plight of destitute respondents to the excessive use of alcohol, while socialist interviewers attributed indigency of their respondents to generally bad economic conditions.

The influence of interviewer expectations on the interview process was demonstrated in 1942 by Stanton and Baker.[6] Respondents were shown 6 geometric designs and were later asked in an interview to select from 12 designs the 6 which they had seen previously. The inter-

viewers were given "inside information" about which designs were originally shown to respondents, but were purposely told the wrong six designs. During the interview, the designs which the interviewers thought were the correct ones were identified more often than the designs originally shown to the respondents.

Another type of study demonstrated a less direct, but still powerful interviewer effect. Katz[7] found in 1942 that interviewers from working-class backgrounds consistently obtained more radical opinions, both social and political, from respondents than did interviewers from the middle class. Robinson and Rohde[8] conducted a study in 1946 on attitudes toward Jews in which interviewers in one group were Jewish in appearance and those in another group appeared to be non-Semitic. The Jewish-appearing interviewers obtained significantly fewer reports of anti-Semitic attitudes than did the interviewers who appeared to be non-Semitic.

Response Error

In a household interview, a respondent can be expected to provide information: (a) that pertains to items about which he is knowledgeable; (b) that he can remember at the time of the interview; and (c) that he is willing to report to an interviewer. Underreporting or inaccuracies in reporting on the part of a respondent may result from lapses in any or all of these three categories.

Myers[9] published data in 1940 from the 1930 decennial census that showed a suspicious pattern of reported ages ending in zero (30, 40, 50, etc.). In the 1930's Twila Neely[10] found that one out of every nine families receiving city relief failed to report this fact. Perry and Crossley[11] published data in 1950 showing that a comparison of interviews with agency records produced significant differences on such items as voting and registration, contributions to the Community Chest, age, and ownership of a library card.

Validity studies comparing data obtained from interviews with data obtained from objective records show discrepancies between the two sources of information for topics such as bank accounts[12,13]; airplane trips[14]; pediatric history[15]; work history[16]; and public health.[3,4,17,18]

One way to interpret underreporting on the part of the respondent is to consider it a consequence of poor memory. The disuse theory, described by Thorndike[19] in 1913 in accordance with the findings in 1885 of Ebbinghaus,[20] suggests that events from the more distant past are more likely to be forgotten than are recent events. Thorndike assumed that the sheer passage of time brings about a weakening of the memory trace. Similarly, one could derive from the Gestalt theory[21,22] a prediction of the high probability of the respondent to forget events of low impact, particularly with the passage of time.

There is, however, another theory of forgetting. According to McGeoch,[23] who first (1932) explicity stated the basic ideas of the modern interferences theory later (1961) expounded by Postman,[24] forgetting does not occur in an absolute sense. Information does not disappear from memory but may be more difficult to retrieve from storage because of competing associations or interferences. Only the accessibility of information declines, resulting in a "lessening probability of retrieval from the storehouse."[25] This would indicate that underreporting is a problem of retrieval, and that reporting can be improved by manipulating conditions that facilitate the recall of information.

Problems of Recall and Information Retrieval

There are two critical stages for a respondent who is asked to report information from memory. First, he has to search for and retrieve the requested information from his memory; then he has to transmit this information to an interviewer. While performance may vary according to the level of the respondent's general motivation or dedication to the role, it is useful to think of recalling and reporting as two specific variables that can affect the accuracy of data. For example, underreporting may result from failure of recall or from failure of communication. An example of the latter case is the tendency of the respondent to withhold threatening or embarrassing information.[26] A fertile field for study is the type of underreport-

ing that results from the failure of the cognitive processes in searching for and retrieving information from memory.

The three major activites of the interviewer are: (a) question asking, (b) probing, and (3) giving feedback. If a question is not properly worded, the probability of obtaining accurate data is low. A question that is improperly worded, inserted out of context, or that conveys to the respondent the type of answer the interviewer wants can produce data that are biased. Probing refers to repetition or rephrasing of a question or the addition of a new question to obtain an adequate response when a previous response has not been adequate. The problem of introducing unwanted bias into the data through probing is solved by distinguishing between directive and nondirective probes. Interviewer feedback consists of evaluative statements that the interviewer makes after the respondent answers a question. These statements may consist of verbalizations indicating approval, attention, or understanding, varying from a simple "Um-hmm," acknowledging the successful communication of an answer, to an elaborate reinforcement of the respondent's behavior.

Several classic experimental studies have demonstrated that simple positive verbal reinforcement can have marked effects on adult performance. Taffel[27] in 1955 gave his experimental subjects a pack of 3-inch-by-5-inch cards each containing a single verb in the simple past tense as well as a list of six pronouns. He instructed each subject to form a sentence from each card beginning with any of the pronouns and using the verb. In the first part of the session, during which the experimente. remained silent, the subject showed a preference for using each of the pronouns on the card. In the second part of the session the task remained the same for the subjects, but the experimenter said "Um-hmm" or "All right" whenever the subjects used either the pronoun "I" or "we" in constructing a sentence. Consequently, the rate of using "I" or "we" increased significantly during the second part of the session.

Research by Greenspoon[28] showed that verbal reinforcement after a respondent mentioned a plural noun in a free-association test increased the rate at which plural nouns were named. Another method of demonstrating the effects of verbal reinforcement involves the occurrence of certain kinds of behavior in a casual conversation setting. Verplanck[29] in 1955 was apparently the first to publish results from this type of study. However, subsequent research[30] indicates that the conditioning effect obtained was probably due to several extraneous variables (primarily experimenter cheating, conscious or unconscious) in using the procedures or reporting the data. More recently, Centers[31] has successfully shown that with great care, one can obtain an increase in the rate with which a person gives opinion statements in a conversation setting if such statements are reinforced by another person.

In 1962, Kanfer and McBrearty[32] interviewed 32 female undergraduates about 4 predetermined topics. During the first part of the interview, in which the women were handed cards designating four topics and asked to talk about each, the experimenter remained silent. During the second part of the session the experimenter reinforced the respondent whenever she talked about a predetermined two of the four possible topics. Reinforcement consisted of a posture of interest, including smiles, and the phrases "I see," "Um-hmm," and "Yes." During the second phase of the experiment the students spent more time talking about the reinforced topics than about those that had not been reinforced.

Bias Introduced Through Interviewer Feedback

The foregoing studies indicate that interviewer feedback may have important effects on the amount of information reported, but they reveal very little about how different kinds of feedback procedures affect interview data.

Hildrum and Brown[33] were the first investigators to show systematically, in a survey interview setting, that interviewer feedback can produce response bias. Two groups of 10 Harvard University students were telephoned and asked their opinions about Harvard Univerisity's philosophy of general education. One group was reinforced by the investigators (who used the word "Good") each time a favorable comment was made, and the other group was reinforced after each unfavorable comment.

Responses of the group reinforced for positive opinions were significantly more favorable toward Harvard's philosophy of general education than those of the group receiving reinforcement for unfavorable comments. Interviewer feedback applied in this systematic way produced a major distortion in the overall attitude responses.

In 1957, Nuthman[34] asked two groups of college students a series of questions about themselves. In one group the experimenter said "Good" when the respondent answered a question in a way that indicated self-acceptance. The other group was given no reinforcement. Respondents who received reinforcement for self-acceptance responses gave more answers of this kind than did the group that was not reinforced.

A. W. Staats and his colleagues have done several studies in this general area. In a 1962 study[35] the experimenter said "Good," "Very good," or "That's fine" whenever the respondent scored in a positive direction on sociability items. Another group was given the same interview but no reinforcement. Staats et al. found that the group receiving the reinforcement scored significantly higher on the sociability scale than did the group not receiving reinforcement. Studies by Singer[36] and Insko[37] followed these general kinds of experimental designs and obtained similar results.

Research Needs

One of the conclusions that can be drawn from this background information on the interview process is that research on the improvement of reporting can fruitfully be devoted to the cause-and-effect relationship between the occurrence of different kinds of behavior or patterns of behavior and the validity of data reported. The behavior that occurs during the interview situation includes not only that of the individual interviewer and respondent, but also that involved in the interaction between the two. Behavior may be motivated by controlled interviewer feedback, techniques designed to facilitate recall, verbal reinforcement, and an effective interviewing instrument, namely, the questionnaire.[38,39]

Obtaining good information in an interview is not simply a matter of asking many questions. More must be learned about the basic principles of memory and retrieval in order to provide a better understanding of the way in which information is stored and to devise more effective ways of retrieving that information.

STUDIES OF UNDERREPORTING OF HEALTH EVENTS IN THE HOUSEHOLD INTERVIEW

This section summarizes some major findings of validity studies about the reporting of health events and health-related behavior in the household interview. It focuses primarily on underreporting, since health events are more likely to be underreported than overreported. Estimates of the magnitude of bias in surveys and calculations of correction indexes for data analysis are not included in this discussion, since the studies show only underreporting bias, not net bias.

The five major studies discussed here were conducted for the National Center for Health Statistics. Their focus was not on the interviewer, but on the characteristics of the respondents and their reporting patterns. Particular attention was also paid to the nature of the information being reported. In five studies similar questionnaires and comparable interviewing procedures were used, and the reports of respondents were compared with independent records assumed to be valid. The studies are identified as follows:

- HIP: a study of the Health Insurance Plan of Greater New York, in which interview reports were compared with medical records;[4]
- SRC: Three studies[1,17,18] conducted by the Survey Research Center, in two of

which reports of hospitalizations were compared with hospital discharge records,[17,18] and a third in which reports of physician visits were compared with clinic records;[1]

- SRI: a study carried out by the Stanford Research Institute in which respondent reports were checked against physician records.[a,3]

Since the studies were designed to investigate validity of reporting and were directed toward problems of underreporting, they were based on samples of records of presumed high accuracy. Hospital discharge records, clinic records, and physicians' records were used as sample frames. Samples were usually weighted for certain characteristics and in some cases certain types of records were omitted from the sample. (For example, in the second SRC study of hospitalization reporting,[18] normal deliveries were omitted and the sample was weighted with hospitalizations of more than 3-months' duration.)

In each of these studies, interviewers were given the family names and addresses of the respondents. Usually a dummy sample was also drawn from the phone book or city directory to help disguise the aims of the study. Interviewers were told that the study was special, but were not told its purpose. Formal inquiry conducted after the studies were completed showed that in no case had an interviewer guessed the study's true purpose.

Interviewers were either experienced in the Health Interview Survey of the National Health Survey, or were given a 2-week intensive training session. Standard interviewing techniques of the Health Interview Survey were used in these studies. While the questionnaires differed in some ways, they were all essentially the same as those used in the Health Interview Survey.

In four of the studies[1,4,17,18] the usual procedure of using proxy respondents was followed. The interviewer personally questioned all adults who were home at the time and used a proxy respondent for all adults not present and for all children. One study (SRI)[3] included only self-respondents.

The analysis consisted of matching reports from the interview with information contained in the medical records. The first part of the analysis involved an examination of the relationship between the characteristics of the health events investigated and the patterns of underreporting. The second part of the analysis was confined to the relationship between characteristics of the respondents and patterns of underreporting.

UNDERREPORTING AND CHARACTERISTICS OF HEALTH EVENTS

Effect of Elapsed Time on Reporting

Investigators have long been aware of the limited timespan over which a person gives accurate reports. However, few studies have had adequate data to demonstrate the extent to which this phenomenon occurs.

Table 1[b] demonstrates the decrease in reporting of hospitalization that occurs as the

Table 1. Number of recorded hospital discharges and percent not reported in interviews, by time elapsed between discharge and interview: Survey Research Center

Time elapsed	Recorded discharges	Percent not reported
1-10 weeks	114	3
11-20 weeks	426	6
21-30 weeks	459	9
31-40 weeks	339	11
41-50 weeks	364	16
51-53 weeks	131	42

Source: reference 17.

[a]A more recent report from this study is *Vital and Health Statistics*, Series 2, No. 57.

[b]Most tables are based on events (hospitalizations, visits to physicians, chronic conditions), not on persons. The person with two events thus has a weight of two. For hospitalizations, 90 percent are single events, and thus persons and events tend to be the same. For physician visits and chronic conditions, however, the concentration is higher.

In tables from the same study, the number of events differs somewhat. For clarity of presentation some irrelevant categories ("not ascertained," for example) are omitted. The full report of these studies gives complete data.

interval increases between the date of the event and the date of the interview. This appears to be a typical "forgetting" curve in which failure to report an event grows as time passes. The same curves are evident for both male and female respondents. The curve rises more slowly for self-reports than for reports given by another family member. Both SRC studies of hospitalization[17, 18] showed very similar patterns.

The HIP study[4] also showed underreporting of hospitalizations (table 2). The numbers are not stable because of the small sample, but the rates of underreporting and the general pattern are similar to those in the SRC studies.[17,18]

Table 2. Number of recorded hospital admissions and percent not reported in interviews, by time elapsed between admission and interview: Health Insurance Plan of Greater New York

Time elapsed	Recorded admissions	Percent not reported
Less than 1-2 months	69	4
2-5 months	82	2
5-8 months	82	4
8-11 months	75	20
10-11 months	42	50

Source: reference 4.

The reader should consider that the interviewing took place over a period of roughly 2 months—from May 2 to July 6, 1958. If the dates of hospital admission are to be expressed as approximate intervals from date of hospital admission to date of household interview, there are overlaps in the classes, but rough equivalents are:

Date of admission to hospital	Approximate interval to household interview
Before July 1957	10-11 months
July-September 1957	8-11 months
October-December 1957	5-8 months
January-March 1958	2-5 months
April-June 1958	Less than 1-2 months

In the SRC study on visits to physicians,[1] respondents were asked to report visits over the 2-week period preceding the week in which the interview took place. As shown below, the rate of underreporting for the second week preceding the interview was twice that for the week immediately preceding the interview.

Time elapsed	Recorded visits	Percent not reported
1 week	196	15
2 weeks	202	30

In the SRI study on reporting of chronic conditions,[3] similar patterns of higher rates of underreporting occurred with an increase in time elapsed since the last clinic visit (table 3).

Table 3. Number of recorded chronic conditions and percent not reported in interviews, by time elapsed between last clinic visit and interview: Stanford Research Institute

Time elapsed	Recorded conditions	Percent not reported
1-7 days	116	9
8-14 days	218	28
15-28 days	440	24
29-56 days	683	42
57-84 days	574	37
85-112 days	513	42
113-140 days	476	45
141-168 days	355	48
169-224 days	372	57
225-280 days	1,232	52
281-364 days	1,078	58
365 days or more	71	59

Source: reference 3.

Similar data for the reporting of chronic conditions for both checklist recognition questions and nonchecklist, free-response items from the HIP study are shown in table 4.

Although the phenomenon of increase in underreporting over time is evident for both hospitalizations and chronic conditions, the shapes of the curves differ. The curve for hospitalization underreporting increases slowly during the 6 months following the event, but increases sharply beyond that period. The curve for the underreporting of chronic conditions rises rapidly during the first few weeks after the visit to the clinic and then flattens out after a

Table 4. Percent of recorded chronic conditions, by checklist status, which were not reported in household interviews, by time elapsed between last clinic visit and interview: Health Insurance Plan of Greater New York

Time elapsed	Percent not reported	
	Conditions on checklist recognition list	Conditions not on checklist recognition list
Less than 2 weeks	32	58
2 weeks-4 months	51	79
4 months or more	66	84

Source: reference 4.

Table 6. Illnesses reported for a 4-week recall period expressed as a percentage of the number reported in the last week of the recall period: California Health Survey

Reported occurrence of illness	Total illnesses	Illnesses with activity restraints or medical attendance	Illnesses without activity restraints or medical attendance
	Percent reported		
Last week	100	100	100
2 weeks ago	60	86	48
3 weeks ago	40	68	28
4 weeks ago	39	66	26

Source: reference 41.

few months. It is interesting to note that data from a feasibility study conducted in Chester, England; Smederevo, Yugoslavia; and Chittenden, Vt.[40] showed that there were significantly fewer visits to physicians reported for the second week preceding the interview than for the first week (see table 5). Similar data on the underreporting of both medically attended and nonmedically attended illnesses over a 4-week reporting period were found in the California Health Survey (see table 6).[41]

Perhaps the best documented phenomenon of underreporting of health events as well as of a wide variety of other types of events and behaviors, is the decrease in the reporting of events as time elapses. This is characteristic of studies of consumer purchases, reports of income, behavior of children as reported by parents, and so forth. Some investigators have hypothesized that this decrease in reporting is not a result of forgetting but is due to the tendency of the respondent to misplace the event in time and recall it as being outside the reference period. This explanation is especially relevant to the sharp increase in underreporting of events of the very last (earliest) weeks of the reporting period. While these studies do not provide a conclusive answer, some of the findings strongly suggest that misplacement in time does not explain a significant amount of underreporting.

Table 5. Percent distribution of reported physician visits, by week of occurrence reported in interview, in three selected areas

Reported occurrence	Chester, England	Smederevo, Yugoslavia	Chittenden, Vermont (U.S.)
	Percent distribution		
Last week	57	53	57
2 weeks ago	43	47	43

Source: reference 40.

In the SRC study of the reporting of physician visits,[1] respondents were asked to report visits made during the 2 weeks preceding the week of the interview. However, the sample was drawn to include persons who had had visits within 4 weeks of the interview. If the underreporting in this study were due to random misplacement of the event in time, one would expect compensatory overreporting of physician visits from the third and fourth week to be reported as having taken place in the first and second week preceding the interview. This did not occur. Telescoping into a more recent time period accounted for only a small amount of the reporting error. In one experimental study in which the usual 12-month reporting period for hospitalizations was lengthened to 18 months, the data were compared to see whether known events were inaccurately reported as having occurred in the 12-18 months prior to the interview. This was not the case.

When respondents who did report their hospitalizations were asked for the month of discharge, 82 percent correctly stated the month and only 3 percent were in error by more than 1

month in either direction. For those who misplaced the month of the hospitalization there was no predominant pattern of reporting the event as having occurred earlier or later. Furthermore, respondents were as accurate in reporting the month of discharge for hospitalizations that occurred between 45 and 52 weeks prior to the interview as they were in reporting those of the most recent weeks. For visits to physicians, over three-quarters of the reported visits were accurately dated to within a day.

These findings present strong evidence that the increase in underreporting as time elapses is not primarily a function of the respondent's inability to place the event in time. One must look to other sources for an adequate explanation.

Effect of Impact of the Event Upon Reporting

Since early studies of memory, it has been recognized that the greater the impact of the event upon the person, the more readily it is recalled. Impact is a term that is poorly defined but generally refers to personal importance or significance of the event. Psychologically, it suggests that certain events occupy a greater part of one's psychic life, having greater relevance than other events for one's present life. In this section some indexes of impact and their relation to underreporting are examined.

Both SRC studies of reporting of hospitalizations[17,18] clearly demonstrate that the longer the duration of the hospitalization, the lower the rate of underreporting. Table 7 shows the results of one of these studies.

Table 7. Number of recorded hospital discharges and percent not reported in interviews, by recorded duration of hospitalization (excluding overreports): Survey Research Center

Recorded duration of hospitalization	Recorded discharges	Percent not reported
1 day	150	26
2-4 days	646	14
5-7 days	456	10
8-14 days	352	10
15-21 days	111	6
22-30 days	58	2
31 days or more	46	8

Source: reference 17.

According to the third SRC study,[1] the level of underreporting of physician visits was lower when two or more such visits had occurred within the 2 weeks prior to the interview:

Recorded individual visits within 2 weeks prior to interview	Total recorded visits	Percent not reported
1	197	29
2	110	21
3 or more	96	13

In the SRI study,[3] a similar decrease in the underreporting of chronic conditions was noted as the number of clinic visits relating to the condition increased (table 8).[3]

Table 9 demonstrates another index of impact. Reporting of automobile accidents was

Table 8. Number of recorded chronic conditions and percent not reported in interviews, by number of individual visits to clinic: Stanford Research Institute

Individual visits	Recorded conditions	Percent not reported
1	3,081	56
2	1,281	47
3	643	35
4	639	26
5 or more	496	14

Source: reference 3.

Table 9. Number of recorded automobile accidents, both involving personal injury and not, and percent not reported in interviews, by time elapsed between accident and interview

Time elapsed	Accidents with no personal injury		Accidents with personal injury	
	Recorded numbers	Percent not reported	Recorded numbers	Percent not reported
Less than 3 months	48	6	71	1
3-6 months	68	12	141	10
6-9 months	48	22	71	10
9-12 months	49	37	94	22

Source: reference 42.

more complete, regardless of the interval since the accident, if personal injury was involved.[42]

Other evidence of the relationship between impact and reporting can be summarized briefly. Hospitalizations that included surgical procedures were more completely reported than those not involving surgery. Conditions are more likely to be reported if the respondent says he has pain and discomfort, is limited in activity, takes medicines or treatment, or is concerned about his health.

Tables 10 (HIP) and 11 (SRC) show the effects of both elapsed time and impact on reporting. The cell totals for the chronic conditions in table 10 are small and the results show some instability, but the previously noted pattern can be observed.

Table 10. Number of recorded service visits and percent of nonchecklist chronic conditions not reported in interviews, by time elapsed between last visit and interview: Health Insurance Plan of Greater New York

Time elapsed	Recorded visits			
	1	2-4	5-9	10 or more
	Percent conditions not reported			
Less than 2 weeks	70	71	25	37
2 weeks-4 months	83	79	60	57
4 months or more	89	85	59	19

Source: reference 4.

Table 11. Recorded duration of hospitalizations and percent of discharges not reported in interviews, by time elapsed between discharge and interview: Survey Research Center

Time elapsed	Duration of hospitalizations		
	1 day	2-4 days	5 days or more
	Percent discharges not reported		
1-20 weeks	21	5	5
21-40 weeks	27	11	7
41-52 weeks	32	34	22

Source: reference 17.

These data suggest that impact level of the event is clearly related to the adequacy of report. Furthermore, there is an interactive effect of impact and time elapsed between the event and the interview. Neither of these relationships is new or surprising; they conform to earlier findings and to theory. What is surprising is the rapidity with which the curve of underreporting rises, especially for chronic conditions, and the strong effect of impact on mediating the effects of time on reporting.

Effect of Social and Personal Threat Upon Reporting

Another factor that affects accuracy of reporting is the level of threat or embarrassment that the requested information holds for the respondent. Much research by social psychologists emphasizes the effectiveness of group norms in bringing about and maintaining approved behavior among group members. Also, one's perceived self-image tends to censor communications so that the image is maintained. The study of hospitalization[17] has some findings on this issue.

A "threat scale" was created for the hospitalization study. The diagnostic classification was a 3-point scale which, in the judgment of the researchers, described the threat or embarrassment involved with the diagnosis. All diagnostic classifications that, in the opinion of the raters, would be very embarrassing or threatening were placed in rank 1. Rank 3 included the groups judged neither embarrassing nor threatening. Rank 2 contained a mixture of categories that were thought to be somewhat threatening or that might be threatening to some persons but not to others. Thus ranks 1 and 3 were kept as pure as possible, and rank 2 contained the uncertain categories. The results of this threat scale, shown in table 12, indicate

Table 12. Number of recorded hospital discharges and percent not reported in interviews, by diagnostic threat rating: Survey Research Center

Diagnostic threat rating	Recorded discharges	Percent not reported
Very threatening	235	21
Somewhat threatening	421	14
Not threatening	1,164	10

Source: reference 17.

Table 13. Number of recorded hospitalizations and percent not reported in interviews, by length of stay, time elapsed between discharge and interview, and diagnostic threat rating: Survey Research Center

Length of stay and time elapsed since discharge	Recorded hospitalizations	Diagnostic threat rating		
		Most threatening	Somewhat threatening	Least threatening
		Percent not reported		
Stay of 1-4 days				
Discharged:				
1-20 weeks ago	223	7	9	7
21-40 weeks ago	355	26	16	9
41-53 weeks ago	219	56	27	27
Stay of 5 days or more				
Discharged:				
1-20 weeks ago	308	0	7	3
21-40 weeks ago	442	15	5	5
41-53 weeks ago	273	33	22	17

Source: reference 17.

that highly threatening or embarrassing information is reported significantly less often than is nonthreatening information.

From table 13, which shows a three-way effect of threat, impact, and time elapsed since hospitalization, it can be seen that there is a low-level relationship between the threat level and completeness of the case of most recent events. For less recent hospitalizations the three factors combine to produce marked differences.

By matching diagnoses from SRC interviews with hospital records two sources of reporting error were found: (a) complete failure to report the hospitalization; and (b) reporting the hospitalizations but misreporting the diagnoses. A few diagnostic categories[43] showing extreme differences between interview data and medical records are examined in table 14. As predicted, those with the lowest reporting levels contain a high proportion of probably threatening diagnoses. There are, of course, reasons other than embarrassment for not reporting a diagnosis accurately; for example, the respondent may not know the diagnosis. However, it is likely that the differences between the two groups are due to differences in threat rather than to other factors.

Undergraduates at The University of Michigan were asked about their hypothetical willingness to report each of a group of diagnostic conditions.[26] In table 15 these data are compared with what respondents actually reported in the HIP study. The diagnostic categories that the students were most willing to report were surprisingly similar to those actually reported best in the HIP study.

Table 14. Number of diagnoses reported in interviews and percent of reported diagnoses compared with hospital records, by selected grouped diagnoses: Survey Research Center

Diagnostic group[1]	Recorded diagnoses	Percent reported compared with hospital records
Benign and unspecified neoplasms	87	+51
Infectious and parasitic diseases	23	+45
Ulcer of stomach and duodenum	36	+12
Diseases of the gall bladder	46	+10
Other digestive system conditions	72	−37
Female breast and genital disorders	52	−44
Diseases of nervous system and sense organs	47	−45
Mental and personality disorders	8	−67

[1]Coded according to *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death*, 1955 revision (World Health Organization, 1957).

Source: reference 17.

Table 15. Hypothetical willingness of students to report certain medical conditions and percent of actual interview reports of these conditions, by medical condition

Condition	Percent willing to report (79 students)	Percent valid reports in household interview (HIP)
More serious conditions:		
Asthma	84	71
Heart disease	58	30
Hernia	55	54
Malignant neoplasm	31	33
Mental disease	19	25
Genitourinary disease	14	22
Less serious conditions:		
Sinusitis	89	48
Indigestion	88	41
Hypertension	83	46
Varicose veins	65	42
Hemorrhoids	21	38

Source: reference 26.

Summary

The data cited here present consistent patterns of reporting; there is a predictable and significant relationship between some characteristics of the information sought and the respondents' reporting behavior. Survey reports are easily susceptible to serious biases in the reporting of health events, and differential reporting bias can result in misleading conclusions. By understanding the problems involved in underreporting and distortion, one can design studies to improve reporting in the interview survey.

UNDERREPORTING AND CHARACTERISTICS OF RESPONDENTS

In the remainder of this section some relationships between reporting and respondent characteristics are analyzed. Are particular respondents most likely to underreport health events? If poor reporting is characteristic of some respondent groups, then the reasons for differential reporting can be examined and experiments can be designed to discover ways to improve reporting. The variables selected for study were those found to differentiate attitudes and behaviors in other studies and which might, therefore, be expected to show differences in the reporting of health events.

Age of Respondent

Data in table 16 suggest that there is an age effect in the reporting of hospitalizations:

Table 16. Number of recorded hospitalization, including and excluding deliveries, and percent not reported in interviews, by age and type of respondent: Survey Research Center

Age of respondent	Recorded hospitalizations	Type of respondent			
		All respondents	Proxy children[1]	Proxy adults	Self-respondents
All hospitalizations		Percent not reported in interview			
18-34 years	792	8	16	13	4
35-54 years	691	10	9	11	11
55 years and over	350	15	–	22	10
Hospitalizations excluding deliveries					
18-34 years	487	12	16	16	6
35-54 years	638	11	9	11	12
55 years and over	349	15	–	22	10

[1] Defined by relationship to head of the household, not by age.

Source: reference 17.

younger respondents showed less underreporting. The apparently large difference in reporting hospitalizations for children (defined by relationship to the head of the household, not by age) is not meaningful since the 35-to-54-year age group reported so few.

Self-respondents tended to be predominantly female and those who have proxy respondents were predominantly male. Since the best reporting was for younger self-respondents, it was hypothesized that this superiority might have been due to the fact that hospitalizations of these respondents might have been heavily weighted with normal birth of babies, a category almost perfectly reported. Deliveries accounted for nearly one-quarter of all the hospitalizations in this first SRC study.[17] The lower part of table 16 shows the underreporting exclusive of normal deliveries.

The overall trend for increased underreporting of hospitalizations with age disappeared when deliveries were excluded. Self-respondents under 35 still showed superior reporting and adults over 55 with proxy respondents showed less accurate reporting. A second study[18] of hospitalization revealed similar age patterns: younger self-respondents showed less underreporting and older persons with proxy respondents more underreporting.

From the third SRC study,[1] a high rate of nonreporting of visits to physicians by respondents 55 years of age and over is shown below:

Age of respondent	Recorded visits	Percent not reported
18-34 years	121	20
35-54 years	200	20
55-74 years	79	34

In the SRI study,[3] in which all persons reported for themselves, chronic conditions were reported more accurately by older respondents (65 years and over) than by younger respondents. This is true for both male and female respondents (see table 17). A second study of the reporting of chronic conditions (HIP) confirms this pattern.[4]

These contrasts in response patterns suggest that the problem of underreporting is not one of memory which usually deteriorates with age.

Table 17. Percent of recorded chronic conditions not reported in interviews, by age and sex of respondent: Stanford Research Institute

Age of respondent	Both sexes	Male	Female
	Percent not reported		
Total, all ages	45	44	46
17-24 years	48	35	57
25-34 years	46	49	45
35-44 years	48	48	48
45-54 years	45	40	49
55-64 years	48	49	47
65-74 years	36	40	32
75-89 years	37	41	31

Source: reference 3.

The seriousness or impact of the condition, the number of conditions, or the frequency of physician visits may influence the reporting level.

The SRI study demonstrated that the number of visits made to physicians during the year was highly correlated with the probability that a known chronic condition would be reported.[3] The nature of the task is another factor which may explain some of the demographic relationships with differential reporting of hospitalization and chronic conditions. For hospitalizations, the respondent was asked whether or not any member of the family had been in the hospital at any time during the past 12 months. For conditions, the respondent was given a list and asked whether or not he had had any of the listed conditions at any time within the past 12 months. Since the conditions were chronic, the probability is high that if a respondent had had any condition at any time during the past year, he would still have had it on the day of the interview. Thus, where both questions appear to ask for recall, the average elapsed timespan was actually much longer for hospitalizations than for chronic conditions. This may explain why the data show a decrease in reporting of hospitalizations over the years, but no similar effect for the reporting of chronic conditions. While these patterns also reflect the effects of other variables, it seems clear that a respondent's age in itself will be a predictor of whether or not health events will be reported.

Sex of Respondent

It has been suggested by some investigators that illness is perceived as some sort of weakness and is more appropriate to the female than to the male role. Admitting to illness may threaten a man's self-image and, therefore, he may underreport illnesses.

Maintenance of the family health is perceived as the role of wife and mother. It can be argued, then, that if one is to use a single respondent to report about the family's health, the wife should be chosen. (This assumes that illness of other family members is not perceived by the woman as a failure in her role performance, which would lead to the prediction of greater underreporting on her part.)

On the reporting of chronic conditions the SRI study[3] showed that males failed to report 44 percent of their own conditions, while females failed to report 46 percent. Similarly, in the HIP study,[4] which compared male and female respondents reporting for themselves or for spouse and children, the reporting difference never exceeded 2 percent.

Male and female respondents showed almost no difference in their reporting of hospitalizations, reporting either about themselves or about other adults or children in the family. Any slight differences were in the direction opposite from that predicted. Similarly, there was no difference between male and female reports of physician visits in either the SRC or the HIP study. In the HIP study, male and female respondents reported as accurately for proxy respondents as for themselves.

One should not make too many generalizations from these results. It must be remembered that the interviewer queried all adults who were at home when she called. A proxy respondent reported for those not at home. Since interviewers usually worked during the day, only people at home during the day were likely to be interviewed. The usual persons at home were housewives, retired or unemployed men, or men who were at home because of illness. The strong possibility existed that these men would be better reporters of health events both for themselves and for others in the family than males who were not interviewed, that is, those who were neither retired nor sick.

Education of Respondent

Since it has been found in some research that persons with more years of education are better respondents, reporting patterns were examined by educational status. The first SRC study of hospitalization showed an interesting pattern: the best reporters were high school and college graduates (table 18). The second hospital study showed the same pattern for high school graduates but the sample size was too small to allow separate consideration of college graduates. Respondents who attended college but did not graduate were poorer reporters than were those in lower educational groups. Whether this pattern is meaningful or is a chance phenomenon is unknown. One could hypothesize that persons who were diligent enough to complete successfully their college educations may also be more diligent in fulfilling demands of other tasks, and thus would be better respondents. Neither study shows a particularly strong tendency for higher educated respondents to report more accurately than respondents with less education.

Table 18. Number of recorded hospitalizations and percent not reported in interviews, by education of respondent: Survey Research Center

Education of respondent	Recorded hospitalizations	Percent not reported
Less than high school graduate	829	13
High school graduate	646	7
Some college	180	16
College graduate	155	5

Source: reference 17.

In the SRC study of reporting physician visits,[1] the high school graduate group did not show the same pattern. As seen in table 19, the college group showed much lower underreporting than did other groups, but the sample was not large enough to warrant any conclusion.

Table 20 shows the underreporting of chronic conditions in the SRI study[3] by educational group. In this study education was reported for the head of the household rather than for the respondent. In this table the pattern of superior reporting with increased education is not apparent. There is no indication here of less

Table 19. Number of recorded physician visits and percent not reported in interviews, by education of respondent: Survey Research Center

Education of respondent	Recorded visits	Percent not reported
0-8 years	121	26
1-3 years high school	132	22
4 years high school	113	23
1 year college or more	33	9

Source: reference 17.

Table 20. Number of recorded chronic conditions and percent not reported in interviews, by education of head of household: Stanford Research Institute

Education of respondent	Recorded conditions	Percent not reported
Less than college	3,983	43
0 years	114	40
1-4 years	110	51
5-8 years	1,151	41
9-12 years	2,608	43
College	2,040	50
1-2 years	851	47
3-4 years	640	53
5 years	549	53

Source: reference 3.

underreporting by high school and college graduates as was found in the hospitalization study.

The one conclusion from these SRI data is that respondents with less than college-level education reported more of their chronic conditions than did those who attended college. In contrast, the HIP study showed no consistent pattern of reporting by educational level.

Data from these studies point to no definite conclusion. One cannot generalize that respondents with more education are better at overall reporting than are those with less education. Why the patterns differ for the studies of hospitalizations and doctor visits from those found in the reporting of chronic conditions is not apparent.

In 1965 Fowler[44] made an intensive analysis of reporting by educational groups in the Health Interview Survey. Based on systematic observation of interviewer and respondent behavior he concluded that less highly educated respondents needed more help from the interviewer to perform adequately. They were less skilled at the respondent role. There was also the tendency for the less educated to have less information about the purpose of the survey and what was being sought in the interview. Interviewers tended to be more active in interviews with less educated respondents, helping them to perform more adequately. Fowler considers that the effect of education may be in the skill level it represents. However, why respondents would show greater skill in reporting hospitalizations than chronic conditions is unclear.

Family Income of Respondent

In other research it has often been found that family income level is a better predictor than either age or education, since income frequently reflects both these variables as well as additional motivational components.

In the first SRC study, hospitalizations were better reported as family income increased (see table 21); in the second SRC study of hospitalization reporting, the same pattern was suggested. The reporting of visits to physicians showed no such trend; although the best reporters seemed to have annual family incomes of $10,000 or more, the sample size was too small to yield firm conclusions.

Table 21. Number of recorded hospitalizations and percent not reported in interviews, by annual family income: Survey Research Center

Family income	Recorded hospitalizations	Percent not reported
Less than $2,000	154	18
$2,000-$3,999	301	13
$4,000-$6,999	760	10
$7,000-$9,999	272	8
$10,000 or more	248	8

Source: reference 17.

In the SRI study on the reporting of chronic conditions, the best reporters were in the lowest income group, with no other pattern apparent (see table 22). The HIP study of chronic condition reporting also showed that persons in families with annual incomes of less than $4,000 were the best reporters and here again, no other pattern was observable. As with education, differences in reporting by income groups are not consistent among studies.

Table 22. Number of recorded chronic conditions and percent not reported in interviews, by annual family income: Stanford Research Institute

Family income	Recorded conditions	Percent not reported
Less than $3,000	639	36
$3,000-$4,999	962	46
$5,000-$6,999	1,373	48
$7,000-$9,999	1,586	46
$10,000 or more	1,437	47

Source: reference 3.

Color of Respondent

The most consistent finding on characteristics of respondents, one which shows up in three out of the four studies, is that white respondents reported significantly better than those of other races (see table 23). This was true whether the respondent was reporting for himself or for other family members. None of the studies involved enough non-white respondents to permit intragroup analysis. One can only hypothesize about the reasons for the finding. On the surface the differences are too large to reflect educational or income factors; rather, they seem to reflect differences in behavior by color.

A similar pattern did not appear in the SRC study of visits to physicians. However, since the sample for that study came from participants in a voluntary health plan, respondents of colors other than white who participated in that plan would be expected to differ in several respects from a random sample.

There is no ready explanation for these reporting differences. It may be that the white interviewer provokes suspicions in respondents of other colors. It may be part of the present cultural pattern for these respondents (especially Negroes) to be unwilling to divulge information. The answers await further experimentation.

Table 23. Percent of recorded hospitalizations, chronic conditions, and physician visits not reported in interviews, by color of respondent

Color of respondent	Hospitalizations		Chronic conditions SRI[3]	Physician visits SRC study 3[4]
	SRC study 1[1]	SRC study 2[2]		
	Percent not reported			
White	10	16	45	24
All others	16	27	50	22

[1]Reference 17.
[2]Reference 18.
[3]Reference 3.
[4]Reference 1.

Reporting for Self Versus Reporting for Other Family Member

In the Health Interview Survey each person at home when the interviewer calls is interviewed for himself, and a "responsible selected adult" reports for persons not at home and for children under 17. As one might expect, the completeness of reporting depends about whom the respondent is talking. It is tempting, in terms of time and cost, to use proxy respondents. However, the data suggest that the practice has some real dangers in terms of quality of responses.

Table 24, covering results of the first SRC study of hospitalization reporting, shows clearly that the more distant the relationship of the respondent to the person about whom information was being reported, the poorer the reporting. The increase in underreporting about children as compared with "self" or "spouse" may be due to the nature of children's hospitalizations, which are generally shorter and involve less serious conditions than those of adults. Data in table 25 from the HIP study on reporting of chronic conditions show a similar pattern of reporting for children.

Table 24. Number of recorded hospitalizations and percent not reported in interviews, by relationship of respondent to sample person: Survey Research Center

Respondent relationship to sample person	Recorded hospitalizations	Percent not reported
Self-respondent	1,092	7
Spouse	275	10
Parent	386	14
Other relation	78	22

Source: reference 17.

Table 25. Percent of recorded conditions, by checklist status, which were not reported in interviews, by relationship of respondent to sample person: Health Insurance Plan of Greater New York

Respondent relationship to sample person	Percent not reported	
	Conditions on checklist	Conditions not on checklist
Self-respondent	57	79
Spouse	62	79
Parent	72	82
Other relation	68	72

Source: reference 4.

Reporting of visits to physicians[1] was better for children than for self-respondents, and was about the same for self-respondents and adults with proxy respondents (table 26). The recall period in this study was only 2 weeks long, and an adult usually accompanied a child to the office; these factors may have accounted for the relatively good reporting of children's visits.[c]

[c]A methodological investigation of the impact of the use of proxy respondents in the Health Interview Survey conducted after the completion of this report is presented in Kovar, M. G., and Wright, R. A., "An Experiment With Alternate Respondent Rules in the National Health Interview Survey," 1975 Social Statistics Section, *Proceedings of the American Statistical Association*, pp. 311-316, and Kovar, M. G., and Wilson, R. W. "Perceived Health Status—How Good Is Proxy Reporting," 1976 Social Statistics Section, *Proceedings of the American Statistical Association*, Vol. II, pp. 495-500.

Table 26. Number of recorded physician visits and percent not reported in interviews, by relationship of respondent to sample person: Survey Research Center

Respondent relationship to sample person	Recorded visits	Percent not reported
Self-respondent	204	25
Parent	103	18
Other relation	96	25

Source: reference 17.

Conclusions

One cannot leave these findings on reporting characteristics of respondents without attempting some explanations. The general picture that emerges from these data is that characteristics of the respondent are not nearly as consistent, nor as strong in their influence on underreporting, as are characteristics of the event.

One finds effects of age, education, and income which are not strong, but which are consistent in the reporting of hospitalizations. The patterns also tend to be consistent for the reporting of chronic conditions. The striking and puzzling fact is the divergent nature of the patterns—persons with higher education, higher income, and of lower ages are better reporters of hospitalizations and poorer reporters of chronic conditions. Although many of the differences are not significant when viewed in isolation, the total impression is that the differences are meaningful and cannot merely be attributed to random error.

It is likely that these patterns reflect the effects of other variables, as has been hypothesized here. The Lansing, Ginsberg, and Bratten study[12] of underreporting of cash loans from loan companies shows a marked income effect, with higher income respondents being poor reporters of their loans. This finding can be understood in terms of social acceptability. Higher income people probably perceive making loans at small loan offices as contrary to the norms of their group. Weiss[45] found that mothers in lower socioeconomic groups are more likely to report that their children were forced to repeat a grade in school than are mothers in higher socioeconomic groups. Again the report may be made to be consistent with

behavior perceived as acceptable. Another explanation may be that lower socioeconomic groups have more sickness; therefore, it has greater impact and is reported better. Hospitalizations, on the other hand, tend to be single events and thus may be more difficult to recall. That the task requirements are different in terms of recall and motivation level are other tenable hypotheses. Research is needed to explain these phenomena.

In the studies presented here there is no indication that special groups are characteristically poor reporters, with the exception of persons of races other than white who are sufficiently consistent in showing high underreporting to suggest that special research be devoted to them.

The general conclusion from these studies is that research on improving reporting can most fruitfully be devoted to the nature of events and the factors underlying the characteristics of events. Problems of elapsed time, impact, and threat or embarrassment appear to be the most significant issues for research.

BEHAVIOR IN INTERVIEWS

Before effective theories about the cause-and-effect sequences in the interview situation can be developed, there must be accurate descriptions and classifications of the material reported in interviews. It is to help meet this need that the Survey Research Center has continued studies which describe the basic nature of the verbal interaction between interviewer and respondent. Since both SRC observation studies discussed here are available in full report form, this discussion will eliminate many of the methodological details and concentrate on the major findings and their possible implications.

HEALTH INTERVIEW SURVEY OBSERVATION STUDY

The first SRC observation study[46] by Cannell, Fowler, and Marquis was carried out in cooperation with the National Center for Health Statistics.[d] Five kinds of measurements were taken for each respondent:

a. Information about respondent demographic characteristics and family health in the regular health interview;
b. A detailed account of the interviewer and respondent behavior as recorded by a third person observing the interview;
c. The interviewer's rating of the respondent following each interview;
d. A reinterview with the respondent conducted by a second interviewer within 2 days following the original health interview; and
e. A staff interview with each health interviewer following the completion of her assignment.

Complete data are available for 412 respondents from a cross section sample of the area east of the Mississippi (excluding the extreme Northeast). About four-fifths of the respondents were women, and about half of the respondents had less than a high school education. Experienced female interviewers employed by the U.S. Bureau of the Census conducted the health interviewing. Another group of women, also employed by the U.S. Bureau of the Census, carried out the behavior observation. The reinterview with respondents was conducted by a Survey Research Center interviewer, rather than the original health interviewer. The staff interview with the health interviewer was also conducted by a Survey Research Center interviewer.

Health Interview Survey Data

The information that the respondent furnished about his own health during the regular health interview was used in this study to create a dependent variable. The dependent variable

[d] A report of this study may be obtained from the National Center for Health Statistics, *Vital and Health Statistics*, PHS Pub. No. 1000-Series 2-No. 26.

was the number of chronic and acute conditions the respondent reported for himself, with adjustment for gross differences in actual sickness which could be predicted from knowing the respondent's age. The previously cited full report of the study details the rationale for the choice of this particular dependent variable. Evidence is presented which suggests that the number of chronic and acute conditions the respondent reports for himself is an indication of the accuracy of other health data reported by him.

Observation

During the health interview an observer recorded what the interviewer and respondent said and did. A wide range of behavior classified in small segments of easily identifiable acts was recorded for both interviewer and respondent. In order to record different kinds of behavior, the interview was divided into segments, each containing a specific set of questions. For each segment several particular kinds of behavior were observed and recorded. In this way, a wide variety of behavior could be recorded while the task was kept within the observer's capabilities.

While the interviewer was still at the door, the observer recorded such things as: the time of day, how long the interviewer had to wait for the respondent to open the door, what the interviewer said as she introduced herself and the study, how many questions the respondent asked, and who took what kind of initiative to get the interview started. The observer also made two ratings about how receptive the respondent had been to this point in the interview. After the actual interviewing started, the observer recorded the occurrence of different kinds of behavior at different points in the interview. Special attention was paid to irrelevant behavior which departed from the task of asking and answering the questions on the questionnaire. Among the categories used to classify this irrelevant behavior were: talking about the other person (such as giving praise), asking irrelevant questions, and giving suggestions. Conversation about the respondent or his family, friends, etc., was also considered irrelevant when it was not directed to the specific question asked. Another major category of irrelevant behavior was humor, consisting of laughter, jokes, and other means of relieving tension. The observer also recorded the reaction which the other person had to each instance of irrelevant behavior. Reactions were rated on a 3-point scale, from "very encouraging" to "very discouraging." Throughout the interview the observer kept track of the kinds of potential distractions present (children, other adults, TV, radio).

During three separate parts of the interview the observer concentrated on the question-answer interaction between the interviewer and the respondent. Seven types of behavior were recorded for the respondent:

a. Adequacy of answer;
b. Elaborateness of response;
c. Inadequacy of answer;
d. Need for clarification or repetition;
e. Checking with another person or with records;
f. Reference to calendar; and
g. Doubt about the adequacy of an answer.

Five specific kinds of interviewer behavior were also counted. They were:

a. Repeating the answer from the questionnaire;
b. Asking a question, not on the questionnaire, which did not suggest an answer (nondirective probe);
c. Asking a question, not from schedule, which might have suggested a specific answer, or asking respondent if she agreed with a specific answer (directive probe);
d. Clarifying the meaning of the question; and
e. Suggesting that records, calendar, or other people be consulted.

Several other attempts were made to examine different aspects of task-oriented behavior. In one section the interviewer counted the number of times the respondent paused before giving an answer, the number of times the respondent asked for clarification or elaborated on an answer, and the number of times the interviewer asked additional questions. During one particularly difficult part of the interview, special attention was given to the frequency with which

the respondent had to ask for help, to the interviewer's behavior when the respondent made such a request, and to the effort made by the respondent during this difficult part of the interview. Between sections of the interview where the observer recorded task-relevant behavior, she recorded her impressions of the respondent's reactions. For example, she rated the respondent's attitude (enthusiastic, bored, irritated), his understanding of the question, and the smoothness of the interaction between interviewer and respondent.

At the end of the interview the observer recorded the length of time spent in conversation after the last question was asked and tried to determine whether the interviewer or the respondent was more willing to continue this conversation. After the interview was completed, the observer filled out two pages of ratings on the respondent and recorded her own impressions of the interview.

Interviewer Ratings of the Respondent

After the health interview, the interviewer rated the respondent by describing her own perceptions of respondent attitude and her own attitude toward the respondent. The rating scales used by the interviewer were similar to the ones used by the observer at the end of each health interview.

Reinterview With the Respondent

A major attempt was made to ascertain the respondent's reactions to being interviewed by conducting a second interview within 2 days following the original health interview. The questionnaire used in this reinterview focused on the respondent's feelings and attitudes about the interview and interviewer: his level of information about surveys in general and this one in particular, his motives for cooperating with the interviewer, and his feelings about the questions and about his role as a respondent. A special attempt was made, using semiprojective techniques, to uncover any negative feelings that the respondent had about the original interview that might be difficult to express directly to another interviewer.

Interview With the Health Interviewer

After all her observed interviews were completed, each of the 35 health interviewers was in turn interviewed by an SRC interviewer. The health interviewer was questioned about her attitudes toward her job, her feelings concerning the interviewing of different kinds of respondents, her reactions to specific aspects of her work, and her reactions to the questions on the health interview schedule.

Results

It was originally hypothesized that several kinds of respondent psychological variables would have a major effect on the quality of data reported. Specifically, it was felt that reporting accuracy would depend on the amount of information which the respondent had about the interview and its sponsors in combination with the respondent's general attitudes, motivation patterns, and particular perceptions of the interview. It was expected that the information level, attitude, motivation, and perception characteristics of the respondent would also be reflected in the behavior observed in the original interview.

This attitude-based interpretation of the causes of accurate and inaccurate reporting is not new. Experience has been accumulated over many years (both from the psychological laboratory and from the world of advertising and marketing) which would enable the researcher to design techniques to change respondent attitudes, motivations, and perceptions and to supply information or correct misinformation. Based on the assumption that poor reporting was due to such variables as low levels of information and inappropriate attitudes, new studies were designed and testing was started on some attitude- and information-level-change techniques. However, as the main data became available, it became clear that the hypotheses concerning the causes of poor reporting were unsupported. There was practically no correlation between the dependent variable (a measure of reporting quality) and the complex indexes of information level, attitudes, motivation, and perception. (See Mueller, Schuessler, and Costner[47] for further information.) These cognitive variables were for the

most part also unrelated to behavior. Some of these data are reproduced in table 27.

As more and more of the data were analyzed, it became apparent that the actual behavior in the interview was the main variable that correlated with the index of reporting quality. Thus it appeared that if changes were to be made in the accuracy with which respondents furnished data about their health, behavior patterns in the conduct of the interview would have to be altered; some kinds of behavior might be more conducive to good reporting than other kinds. To test this, correlations were obtained between the reporting index and the frequency of various kinds of behavior which took place during the interview. The preliminary results suggested that the kinds of behavior normally considered task oriented (asking for clarification, giving elaborations upon answers, and consulting records, on the part of the respondent, and probing by the interviewer) were more highly correlated with the dependent variable than the kinds of behavior which are considered to be less relevant to task performance, such as talking about self or joking. To illustrate, the relationship between respondent behavior and reporting[31]:

Table 27. Relationship of respondent cognitive variables to reporting index and behavior indexes

Cognitive variables	Gamma coefficients of association[1]		
	Reporting index	Behavior index	
		Task-oriented	Irrelevant
General feelings toward interview			
Direct questions	.01	.01	.06
Semiprojective questions	.03	.06	.07
Stated reasons for cooperating			
Citizen's duty	.08	.03	.00
Desire to talk	−.05	.02	.12
Personal benefit	−.13	.21	.14
Opportunity for a break in routine	.08	−.06	.20
Concern about health	−.07	−.14	−.20
Stated reasons for not wanting to cooperate			
Concern about time	−.01	.00	.00
Concern about the questions	.15	.12	.08
Perceptions of the respondent task			
Interviewer wanted exact answers	.18	.07	.06
Interviewer wanted everything to be reported	.10	−.13	−.02
Reason for collection of information (statistics)	.22	.13	.02

[1]None of the coefficients of association is significantly different from zero at the 5-percent level. Gamma is a nonparametric measure of association based on rank order that ranges from −1 to +1. Near zero, it indicates little association between the variables being tested. For further information see reference 47.

Respondent behavior index	Gamma, reporting index[1]
Task-oriented	.56
Interpersonal	.22

[1]Both coefficients are significantly different from zero ($p < .05$).

Further investigation revealed, however, that the frequency with which any one category of behavior occurred in the interview was highly correlated with the frequency with which any other kind of behavior occurred. Thus it was not found that some interviews were predominantly task oriented and others predominantly interpersonally oriented. What was found was a general behavior activity level characterizing a particular interview. The higher the behavior level (the more frequently each kind of behavior occurred for both interviewer and respondent), the higher the score on the reporting index.

Since it was impossible to determine which of the behavior categories, if any, determined this general activity level, it appeared logical to ascertain who was responsible for setting the activity levels. Since the data are correlational, it

is difficult to determine directly whether the respondent or the interviewer had major responsibility for determining the amount or level of behavior in the interview. However, it was determined that interviewers themselves did not have a characteristic behavior level for all interviews. The data also indicated that there was an extremely high correlation between the level of behavior of the interviewer and the level of behavior of the respondent:

Interviewer behavior index	Respondent behavior index[1]	
	Task-oriented	Interpersonal
Task-oriented	.84	...
Interpersonal	...	.85

[1]Both gamma coefficients are significantly different from zero ($p < .05$).

Thus it appeared that the amount of behavior in the interview tended toward some sort of balance. If the interviewer engaged in a high level of behavior, so did the respondent, and vice versa. It was also noted that the balance was most likely to occur when the behavior levels of the interviewer and respondent were either especially high or especially low. A special statistical treatment of the behavior index data which shows the high probability of balance at extreme behavior levels is given in the original research report by Cannell, Fowler, and Marquis,[46] pp. 26-27.

The major conclusion from this observation study was that the original hypotheses about the effects of such variables as information and attitudes on the quality of reporting were probably wrong. The major causes of good and bad reporting are probably to be found within the interview itself, particularly in the behavior of the participants. It was unclear which variables determined the behavior of participants. The interviewer could be responsible for setting the behavior level, the respondent could have primary responsibility, both could share equal responsibility, or the behavior level could be determined by some other variable or variables. The data led to speculation on a procedure referred to as a "cue search" model of interview interaction. It may be that the household interview is a rather unique experience for the respondent and that he really does not have a set of predetermined behavior patterns for it. The newness of the interview situation might make it difficult to generalize his associated feelings, attitudes, and expectations. The respondent must look to the interviewer or to some other source for cues about expected behavior. On the other hand, the interviewer may be in somewhat the same situation. She has learned from experience that respondents are different: some will enjoy the interview and others will be annoyed by it, some will have trouble with certain sections of the questionnaire while others will not. Therefore, the interviewer will be attentive to subtle cues from the respondent to help her arrive at a strategy for dealing with each particular interview.

This hypothesis implies that both interviewer and respondent search for cues from each other about appropriate kinds of behavior. This cue-searching process could account for the strong tendency of interviewer and respondent to behave at the same level of activity in the interview. This heavy reliance on cues from the other person to set the behavior pattern may also account for the fact that cognitive orientations measured in this study were not predictive of behavior or reporting level. In addition, the reciprocal cue-searching process may explain why this research did not determine whether one person sets the behavior activity level and the other follows.

Subsequent research has shown that changing the characteristics of interviewer behavior can have marked effects on both the amount and the accuracy of health data reported by respondents. These studies, while quite limited, support the general interpretation of the findings of the first observation study: namely, that changes in response accuracy are most likely to be achieved by changing the interaction process in the interview itself. These studies also show that changes in interviewer behavior will often be accompanied by changes in both respondent behavior and reporting accuracy. They do not rule out the possibility that data accuracy may be significantly affected by the respondent and other sources of variation, but they show that the interviewer can have at least some beneficial effect independent of other possible influences.

URBAN EMPLOYMENT SURVEY BEHAVIOR INTERACTION STUDY

With the cooperation of the U.S. Department of Labor, Urban Employment Survey, the Survey Research Center conducted another study[2] of the behavior of the interviewer and respondent in the household interview. This study by Marquis and Cannell differed in a number of ways from the observation study described above. Data on the verbal behavior of the interviewer and respondent were obtained through a tape recording of the interview rather than through the recording of impressions by a third-person observer. The tape-recording procedure substantially reduces data collection costs and allows a much more detailed and refined coding of the verbal behavior that occurs during the interview.

In this study, a cross-sectional sample of 181 employed male respondents residing within the city limits of Detroit were interviewed. There were four interviewers—all of whom were white, female, middle-aged, and residents of the suburbs.

New Coding Scheme

The study employed a revised coding scheme for interviewer and respondent verbal behavior. The coding scheme omitted all nonverbal behavior. It also included more code categories for task-related behavior: several categories reflected the way in which the question was asked, and more detailed codes recorded the way the respondent answered questions. Because of the research on the effects of interviewer reinforcement which had taken place between the first and second observation studies, a code for interviewer feedback and respondent feedback was included in the new scheme. Several codes dealing with irrelevant conversation were deleted since they had not proved useful in the previous study. A summary of the new coding scheme is presented in table 28. Additional items derived from more recent studies end the table.

Main Findings

Behavior balance.—One of the main findings from the original observation study was that the behavior of the interviewer and the respondent was best described in terms of a "balance" model. That is, if one person was engaging in a great deal of behavior, so was the other. This is in contrast to another pattern which might be expected: namely, that a low level of respondent behavior would be compensated for by a high level of interviewer behavior, and conversely, a high level of respondent behavior would be accompanied by a low level of interviewer behavior. In this second observation study it was possible to control, both in the questionnaire design and in the statistical treatment of the data, the number of questions asked over the entire interview. Thus it was possible to compute an index of interviewer behavior level and respondent behavior level per question. The ability to control the number of questions made it possible to remove one source of variation which might have accounted for the behavior balance phenomenon observed in the first observation study. The correlation coefficient between the amount of interview behavior per question and the amount of respondent behavior per question was .77. This demonstrates again the strong interdependence of interviewer and respondent behavior during the interview. It also indicates that the variables or parameters which have a causal effect on reporting quality are probably to be found in the behavior interaction within the interview rather than in the personal characteristics (e.g., attitudes) of either of the participants.

Question asking and probing.—Because of the expanded coding scheme and because the interactions were tape recorded rather than coded during the interview, the second study provided a much more detailed description of the kinds of verbal behavior that occurred during the interview. The descriptive data confirmed that those interviewing procedures for which the interviewers were trained, such as question asking and probing, were carried out effectively in accordance with accepted procedures. Interviewers asked the question in the correct manner more than 90 percent of the time. Respondents gave many answers which did not meet the objectives of the question, and interviewers used many probes in attempts to get adequate information. The probes used were generally nondirective; that is, they did not

Table 28. Summary of revised coding scheme for interviewer and respondent verbal behavior used in the urban employment survey behavior interaction study

Code item	Description
	Interviewer behavior
Correct question	Question asked essentially as written on the questionnaire
Incomplete question	Part of a question correct as far as it goes
Inappropriate question	Question that was asked but should have been skipped due to a skip pattern
Incorrect question	Question in which meaningful word(s) have been altered or omitted
Repeat question	Question that has already been asked and is asked correctly again
Omitted question	Question omitted by mistake, contrary to the questionnaire instructions, and for which the relevant information has not been obtained by means of a preceding question
Information previously obtained	Question omitted because sufficient information to code an adequate response has previously been volunteered by the respondent in answer to a prior question
Skip pattern	Question omitted because of skip pattern prescribed by the questionnaire instructions
Nondirective probe[1]	A probe that neither suggests a specific answer or class of answers nor restricts the frame of reference of the original question
Directive probe[1]	A probe that suggests possible responses or implies that some answers are more acceptable than others. It restricts the frame of reference of the original question.
Gives clarification	Gives clarification upon request of the respondent regardless of whether the information supplied is correct or incorrect. Includes also rephrasing or explanations of questions.
Volunteers information	Volunteers information relevant to the topic of the question or interview. Includes transition statements.
	Respondent behavior
Adequate answer	An adequate response to a correctly asked question that meets the objectives of the question as stated in the Interviewer's Manual. Incorrect clarification does not rule out the occurrence of an adequate answer. May also occur as the result of a probe, provided the response meets the question objective.
Inadequate answer	An inadequate response to a properly asked question that does not meet the question objectives as stated in the Interviewer's Manual
Don't know answer	Response to a correct question that indicates that the respondent does not know, only if *not* followed by an attempt to answer the question
Refuses answer	Verbal refusal to answer question
Other answer	Response (to an incomplete or incorrect question or a response to a probe) that does not meet the question objective
Elaboration	Gives reason for a response or supplies more information that required for an adequate answer and is relevant to the question topic
Asks clarification	Requests clarification of a question or question objective

See footnotes at end of table.

Table 28. Summary of revised coding scheme for interviewer and respondent verbal behavior used in the urban employment survey behavior interaction study—Con.

Code item	Description
	Behavior of both interviewer and respondent
Feedback	Behavior that indicates attention, approval, understanding, or how well the other person is doing, only if not a response to a question or a probe (excluding "Thank you")
Ongoing feedback	Ongoing feedback that indicates attention, approval, understanding, or a desire to interrupt while the other is talking *without* successfully interrupting the speech of the other person
Repeats answer	Repetition of response either exactly or as a summary, or utilization of previous responses for transition to a new topic or for asking a question or a probe
Irrelevant conversation	Statements unrelated to the question or general field of the inquiry. Generally rapport-building or personal rather than task-oriented behaviors.
Gives suggestion	Suggests new kind of behavior that will enhance, interrupt, or resume task behavior
Polite behavior	Polite behavior or socially expected courtesies not specifically related to task and not included in the printed question on the questionnaire (e.g., "Please," "Thank you")
Interruption	Successful interruption. The other person must stop talking. Blocks can't occur at the end of a sentence or at a pause which might be considered the end of a question.
Laugh	Audible laugh, chuckle, or snicker that may indicate humor, tension, or ridicule
Other	Any significant behavior not elsewhere coded or unintelligible verbal behavior
Extraneous interaction	Interaction of either interviewer or respondent with a third person during the interview
	ADDITIONAL CATEGORIES[2]
	Interviewer behavior
Modified question	Question worded *essentially* as written but with unimportant modifications
Alternatives incomplete	Fails to read all or some of response alternatives *because* interrupted by respondent
Infers answer	Omits question because interviewer can infer the answer even though it has never been stated explicitly by respondent
"Anything else" probe	Special case of nondirective probe
Invents task-oriented question	Invents new question to gain or confirm information necessary to follow skip instructions
	Respondent behavior
Additional response	Given for each adequate answer beyond the first one for open-ended questions
Closure	Respondent indicates no more information to give on open-ended question
Simultaneous answer	Respondent answers while interviewer is talking

[1] A probe is a question or statement used by the interviewer to elicit further information. It is a creation of the interviewer and is not on the questionnaire.

[2] Additional categories have been derived from more recent studies.

suggest to the respondent any particular answers. This kind of probing, according to the theory, helps to avoid the introduction of interviewer bias. Furthermore, the data suggest that the probing theory, with its emphasis on nondirective as opposed to directive probing, is correct. In this study interviewers were much more successful in getting adequate answers after nondirective probes than after directive probes. These results are tentative however, since it is possible that interviewers used nondirective probes when they expected that the respondent would have no trouble in giving an adequate answer and used directive probes only when they anticipated a great deal of difficulty in getting an adequate answer. Natural observation studies of this type are subject to this limitation on inference. Experimental studies are needed for a more refined analysis of the actual cause-and-effect relationships.

Feedback and nonprogramed behavior.— Another major finding from these data is that a large proportion of interviewer and respondent behavior is nonprogramed behavior. That is, much that goes on during a household interview is not considered in typical interviewer training. Two sets of data illustrate this point.

One discovery was that interviewer feedback, the interviewer's verbal reaction to the respondent's answer (such as "O.K.," "I see," "Good,"), occurs very frequently. In fact, in this study interviewer feedback accounted for about 23 percent of all interviewer behavior coded. The effect which interviewer feedback may have on the accuracy of these data is discussed in detail in the section "Use of Feedback To Increase Accuracy," of this report. These observational data indicate that feedback is very frequent and is probably used in a way that is nonproductive, or even counterproductive, of good data. Specifically, the data indicate that positive feedback statements occurred just as frequently after inadequate answers as after adequate answers. Most surprising was the finding that positive feedback statements were used over half the time when the respondent refused to answer a question. The probability of the interviewer using a feedback statement after nine different kinds of respondent behavior is shown in table 29.

Table 29. Probability of interviewer feedback following respondent behavior, by category of respondent behavior

Category of respondent behavior	Probability that interviewer feedback follows
Adequate answer	.28
Inadequate answer	.24
"Don't know" answer	.18
Refusal to answer	.55
Other answer	.34
Elaboration	.30
Repeats answer	.32
Gives suggestion	.33
Other behavior (not classified elsewhere)	.21

The significance of the pattern of feedback use demonstrated in this study is not entirely clear, but results from this and other studies lead to some hypotheses. In another section of this report several experimental studies are described in which different ways of using positive feedback statements were tested (see "Use of Feedback in a Nonexperimental Interview," this report). These studies show that if positive feedback statements were used by the interviewer only after the respondent has given an adequate answer, more accurate data were reported than when no feedback statements were used. The data in table 29 indicate that the interviewers used feedback in a random fashion or when they felt some tension was developing or about to develop. Neither the effects of random feedback contingencies nor of tension-reduction contingencies have been evaluated in the household interview setting. An educated guess at this point is that these strategies are less productive of accurate reporting than the usual laboratory strategy which provides verbal reinforcement only after desired respondent behavior. Further research is planned in this area.

Another technique for determining if there is more to the personal interview than asking questions and giving answers is to divide the behavior data into two parts: (a) the average amount of behavior needed to get an adequate answer to a question, and (b) the average

amount of behavior which occurs after an adequate answer and before the next question. Data in this study show that, on the average, one-third of all behavior occurred after an adequate answer and before the beginning of the next question. Computer programs are being modified to explore in greater detail the kind of behavior that takes place after an adequate answer. Although these results are not yet available, it is possible that this "extra" behavior may represent a large potential either for bias in the data or for cues which lead to even more accurate information.

Effect of type of question on behavior.—The behavioral data from this study make it possible to explore how different kinds of questions result in different kinds of behavior patterns. In his doctoral thesis, Thomas deKoning[48] classified questions on two independent dimensions: open-closed and fact-attitude. An open question was defined as any question to which the respondent must formulate his own answer, while a closed question was defined as one to which the respondent might either anwer "Yes" or "No," or respond according to the alternatives supplied in the question. DeKoning's closed question was similar to what others call a forced-choice question. The dependent variable was the average number of behavior codes assigned per question. Results which are summarized in table 30 indicate that, as might be expected, there was more behavior recorded for open questions than for closed questions. When these data are split into specific behavior categories, it appears that interviewers probe about three times as often and provide about twice as much feedback for open as compared with closed questions. On the other hand, the respondent is about six times as likely to give an unacceptable answer to an open question as to a closed question. This pattern of results suggests that open questions present the respondent with more difficulty in meeting the objectives of the question. This conclusion is supported by data presented by "Question Length and Reporting Behavior," in another section of this report, and by the results of an experimental study by Marshall, Marquis, and Oskamp[49]. Open questions require the respondent to retrieve information from memory with a minimum of stimulating cues. On the other hand, closed questions involve only recognition memory.

Table 30. Rate of recorded behavior for open and closed questions, by type of behavior

Type of behavior	Rate of behavior per question	
	Open questions	Closed questions
Interviewer behavior		
Total	3.57	2.13
Correct question	0.91	0.91
Probing	1.05	0.32
Feedback	0.86	0.46
Other behavior	0.75	0.44
Respondent behavior		
Total	2.76	1.68
Adequate answer	.84	.86
Unacceptable answer (inadequate, don't know, refusal)	.36	.06
Other answer	.63	.35
Elaboration	.39	.24
Other behavior	.54	.37

That is, all the respondent is required to do in response to a closed question is to decide whether the stated characteristic is true or false, good or bad. While such questions do have advantages of clarity and ease of recall, it is often necessary to ask many of them in order to cover the same material as is covered by one open-ended question.

DeKoning[48] also showed that there is a higher behavioral level in getting an answer to an attitude question than to a question of fact. However, the differences are not quite so large as for open and closed questions. For attitude questions interviewers are more likely to probe, to give feedback, and to engage in irrelevant behavior and laughter than for fact questions. Respondents are more likely to give unacceptable answers, to ask for clarification, and to elaborate upon their answers when responding to attitude questions. These data suggest that attitude questions are somewhat more difficult and cumbersome to handle than are fact questions, but this difference is small and may

be due to other variables confounded with the attitude-fact distinction.

Diagnosing specific question problems.—One of the intriguing motives for using the behavior coding technique is to arrive at a systematic evaluation of the adequacy of specific questions as they appear on the interview schedule. This kind of evaluation procedure may be very useful in the pretest phase of questionnaire construction. Social science strives to be a scientific discipline, but the procedures used by social scientists to develop and validate questions and questionnaires are generally crude. One usually sends a group of interviewers into the field with the questionnaire developed in the office. There is then a meeting (or series of meetings) during which interviewers and the researcher discuss the questionnaire. One hears familiar statements such as "This question seems to work well," or "This question seems to do what we want it to because we have the distribution of responses." The interviewer might say, "I don't think the respondents really understood this question," or "This question irritates people." It is on the basis of such highly subjective evaluations that questionnaires are developed.

In preliminary work done for the U.S. Department of Labor, the behavior-coding method has shown considerable promise for evaluating some aspects of pretest interviewing.

There seem to be three kinds of problems with questions: (a) those attributable to the interviewer, (b) those that reflect respondent difficulty, such as failure to understand the question or trouble in recalling information, and (c) problems caused by the questions themselves, such as poor syntax, "skip" instructions that are difficult to follow, or obscure placement on the interview schedule. A small-scale attempt was made to trace question problems to one or more of these sources by using a small number of the codes obtained in the behavior observation study.

For example, those questions that were asked incorrectly by the interviewers at least 15 percent of the time were identified. A question was coded as "incorrectly asked" if important words or phrases were changed or omitted.[e] The list of incorrectly asked questions revealed several items which contained parenthetical phrases, others which contained difficult syntax, and still others which were extremely cumbersome to handle in verbal form. The first set of questions pointed to the fact that during training, interviewers had been given inconsistent rules about handling parenthetical phrases. The second group of questions, containing awkward syntax, pointed to a problem that has been overlooked by many questionnaire designers: When questions are extremely long and complex, respondents often interrupt at the end of a clause to answer without allowing the interviewer to finish the question.

Another set of 39 questions was identified as having been answered inadequately more than 14 percent of the time even though they were asked correctly. From this list there appeared to be two reasons why a question would receive an inadequate answer code a high percentage of the time: (a) the respondent was unable to answer because the required information was not easily accessible from memory, and (b) the interviewer could not discriminate between an adequate and inadequate answer, and therefore mistakenly accepted the inadequate answer as meeting the objective of the question.

Another analysis of question problems was tried in which two kinds of interviewer omission codes were combined to show different kinds of question problems. The logic of that analysis is as follows:

Nature of problem	Code (N)[1]	Code (*)[2]
Interviewer error	High	High
Questionnaire redundancy	High	Low
Skip pattern or format problem	Low	High
No problem	Low	Low

[1]Code N indicates a question was omitted because the answer was already given.

[2]Code * indicates that the question was omitted by mistake.

NOTES: "High" indicates the question was omitted 10 percent of the time or more.

"Low" indicates the question was omitted less than 10 percent of the time.

There were 27 questions identified as being omitted many times, either because of error or

[e]More stringent criteria were used in the last section of the questionnaire which contained mostly attitude questions.

because the interviewer thought the answer had already been given. The data suggest that omission problems like these might be overcome if interviewers received better instruction as to what constitutes an adequate answer. Thirteen questions were identified as belonging to group three, questions which were omitted often by mistake but were not skipped because an answer had already been obtained. This finding also suggests the need for better interviewer training since these questions are often skipped because the interviewers assume they have been answered adequately through previous questions—when indeed they have not.

On the other hand, better interviewer training concerning the objectives of each question may not entirely solve the omission problem. Other data indicate that omission rates are above 10 percent only when a question is to be asked of a subsample of respondents. Questions which must be asked of all respondents are almost never subject to high omission rates. Of the 71 questions in this interview which were to be asked of all respondents, only 1 was omitted more than 10 percent of the time, while of the 102 questions to be asked of subsamples of respondents, 55 (54 percent) were skipped more than 10 percent of the time, either by mistake or because an adequate answer had already been obtained. Thus, omission problems may be traced to skip instructions and other subsampling techniques. While these procedures are often necessary, the questionnaire designer should be aware of the potential for interviewer omission error whenever subsampling techniques are used. The subsampling omission bias may be especially acute in an interview such as the one tested, in which skip patterns occur frequently.

Other procedures to identify question problems have been or will be tried. The possibility of obtaining systematic data on question problems in the pretest phase of a survey study remains intriguing. Much work is still to be done in devising procedures relating to question design. The Survey Research Center methodology program is working to develop additional kinds of logical analysis of question problems, as well as to reduce the cost and time involved in obtaining such data. When some of these latter problems are solved, other survey research agencies should be encouraged to experiment with these appropriate procedures. There are enough experimental and observation studies now in the literature to indicate that question variables such as structure and content are important determinants of data accuracy in survey interviews. While many factors may potentially affect data accuracy, the question variables probably have a much greater potential effect on overall accuracy and completeness than does any other single class of variables. A thorough understanding of how question construction and question content affect data accuracy should do a great deal to advance the usefulness of the survey interview for research purposes.

Effect of respondent age and race on behavior.—Social scientists and those responsible for the conduct of cross-sectional sample surveys often hypothesize that the demographic characteristics of the interviewer and the respondent, such as age, race, education, and income, will have some effect on survey data accuracy. For example, earlier research has shown that white respondents are reluctant to admit prejudice toward blacks when the interviewer is black. Such results are interpreted as being reasonable in terms of cultural norms concerning prejudice.

The question remains, however, whether the results of these studies are applicable to the reporting of all information in all survey interviews. For example, there is evidence to suggest that the accuracy of data obtained about such things as physician visits or financial facts does not differ by race of respondent. This second SRC observation study included an experimental design which provided for the investigation of interaction patterns by respondent age and race. White middle-class female interviewers interviewed employed male respondents. There were four experimental groups of respondents: (a) white, 18-34 years (N=47); (b) black, 18-34 years (N=44); (c) white, 35-64 years (N=43); and (d) black, 35-64 years (N=47).

With respect to the effect of age on behavior during the interview, the data confirm the results of the first observation study. Older respondents were much more likely to engage in a large amount and wide variety of behavior

during the interview. The proportion of their behavior devoted to good task performance was much lower than that of younger respondents. In addition, when interviewing older respondents, the interviewers displayed high frequencies of a variety of behavior. Thus, an interview with a younger respondent was quite different from an interview with an older respondent. The former appeared to be task oriented while the latter was characterized by a great deal of extra behavior which may have resulted in keeping the interaction at a relatively tension-free level.

The effects of respondent race on the kind of behavior shown in the interview were not clear. Two attempts have been made to analyze and explain these data,[2,48] and both produced a somewhat similar set of inferences based for the most part on nonsignificant statistical trends. The overriding conclusion is that when age was controlled, the effect of respondent race on the kind of behavior in an interview was not marked for the kind of information contained in the interview. However, the data do suggest that if there was a race effect, it was in the areas of interviewer probing behavior and respondent inadequate answering behavior. Although the differences were not always statistically significant, it appeared that the proportion of inadequate answers (answers which normally require probing) was higher among black than among white respondents, and that interviewers probed more with black than with white respondents. Also, black respondents tended to give more "don't know" answers, repeat more of their answers, and ask for clarification more frequently than did their white counterparts. These racial differences are very small but may indicate slight differences in difficulty with the questions. The pattern viewed as a whole does not indicate any active resistance or lack of motivation to cooperate.

There was a slight tendency for white respondents to exhibit more ability to give adequate answers than did their black counterparts, and interactions with the female interviewers seemed to be less task oriented among white respondents. For example, white male respondents engaged in slightly more polite behavior, feedback, and elaborations than did black males. White males also seemed to show more resistance or more dominance while performing their task, as indicated by a slightly higher percentage of refusals to answer, suggestions to the interviewer, and unsuccessful attempts to interrupt the interviewer.

In summary, the age effect was found to be fairly reliable. It was quantitative rather than qualitative. Older respondents in comparison to younger respondents engaged in a higher percentage of almost every kind of behavior except providing adequate answers and making requests for clarification. The race effect was much smaller than was the age effect. Black respondents showed a pattern of behavior characteristic of well-motivated performance on a difficult task. White respondents seemed to have an easier time at the task, interacted more smoothly with the interviewer, and showed a slightly greater tendency toward dominance or resistance. It seems likely that whatever differences exist may reflect variables such as educational background of respondents rather than race as such.

Interpretation problems—The difficulty in trying to interpret the nonsignificant differences between the two racial groups points to an apparent problem or handicap in the current behavior observation scheme. The problem seems to be that the readily coded behavior categories such as "asks question correctly," "refuses to answer," and "laughs" are difficult to define in an abstract sense. Social scientists are accustomed to dealing with abstract concepts about human interaction, such as: "shows hostility," "is annoyed," "interacts smoothly," or "is having conceptual difficulty." At an even higher level of abstraction these concepts might be: "is ingratiating," "shows lack of rapport," "enjoys the interview," or "is motivated." A point to be made in defense of the observation technique and its scheme of categorizing behavior into small units is that extrapolating behavior codes to a higher level of abstraction does not really provide much more meaning to the data. The theories of human interaction to which some have attempted to fit the existing data have not themselves been validated to any great extent. Interpreting the present data in these frames of reference will

probably not yield any greater understanding or predictive power.

It is suggested, however, that the problem of attributing meaning to the observational data may be carried out in a different way. If it is recognized that the problem of the survey researcher is to obtain complete and valid data, the strategy for assigning meaning to the various behavior observation codes becomes fairly clear. Empirical research is needed to establish the cause-and-effect relationship between the occurrence of different kinds of behavior or patterns of behavior and the validity of data reported. Thus, the next research steps might include several kinds of studies which record and code the behavior that takes place in an interview and which, in addition, obtain independent verification of the accuracy and completeness of the data the respondent has reported in his answers. Correlations between the accuracy measures and the behavior measures would then yield significant insights into the meaning of the behavior codes or combinations of behavior codes.

INTERVIEWER PERFORMANCE DIFFERENCE: SOME IMPLICATIONS FOR FIELD SUPERVISION AND TRAINING

Most survey researchers believe the adage that practice makes perfect, or at least makes for improvement. Thus, they expect seasoned interviewers when compared with less experienced ones to be more skilled at adapting to various interviewing situations, more at ease in interacting with respondents from various social classes, and more proficient in using nonbiasing procedures. Interviews conducted by an experienced staff usually present fewer problems in coding; the responses are clearer and more adequate to the objectives, contingencies are followed more strictly, and, with a minimum of omitted questions, there are fewer noncodable replies.

Data from methodological studies which were recently examined raise questions about the positive effects of experience in interviewing. Since the analysis of interviewer behavior was not planned as part of these studies, the research designs are inadequate to produce conclusive findings but are sufficiently intriguing to encourage further study. As an incidental analysis in one study, the data on failure to report known hospitalizations were tabulated for each item separately. These data showed a surprising trend: The larger the number of interviews taken by a single interviewer, the fewer the hospitalizations reported by the respondents. Although random assignment of interviewers was not made, and the results therefore might reflect a difference in types of respondents interviewed by each interviewer, the finding was sufficiently interesting to raise questions about the positive effects of experience in interviewing and to lead to an analysis of interviewer performance in other studies.

The first SRC study[17] was designed to investigate the underreporting of hospitalizations. A sample of approximately 2,000 hospital records was selected from patients who were discharged from a hospital within the 12 months preceding the interview. The sample was taken from hospitals located in counties that were a part of the Health Interview Survey (HIS)[f] regular national sample. The family name and address of the person discharged were given to the Bureau of the Census interviewers who regularly conducted the interviews for the interview survey in that county. Twenty-seven experienced Bureau of the Census interviewers were included in this study. All interviewers working in the areas in which sample hospitals were located did some interviewing. Because the number of sampled discharges varied considerably by county, interviewers received varying numbers of sample addresses each week.

The questionnaire and interviewing procedures used were identical to those of the regular

[f] For details of the sample, the procedures used, and the findings of this Health Interview Survey, see reference 17.

Health Interview Survey. Interviewers were informed by the census field office that they were to undertake a special study and were given special sampling instructions. They were told that the regular questionnaire and interviewing procedures were to be followed. Interviews were conducted with each adult member of the family found at home at the time of the visit. For those not at home and for all children, a proxy "responsible adult" was interviewed.

While the interviewers were aware that this was a special study, the purpose was not divulged. In order to attenuate the number of hospitalizations reported, a sample of names and addresses drawn from a telephone directory was added. However, interviews at these addresses were not used in the analysis.

Following the usual procedure, each interviewer was given a weekly assignment which she was to complete during that week. The interviewing extended over 3 months, and the interview reports were matched with the hospital discharge records. Table 31 shows the rate at which hospitalizations were underreported[4] by the number of interviews taken by groups of interviewers. The data show a tendency for the rate of underreporting of hospitalization to increase as the number of interviews increases. The rank order correlation of the number of interviews taken and the failure to report the hospitalization is very high.

Attempts to understand these results by looking for differences in characteristics of hospitalizations were not fruitful. The overall response rate for this study was 95 percent; thus differences could not be attributed to low response. Interviewers were given a weekly assignment depending on the number of sample discharges in their county or that part of the county in which they worked; thus they had no choice in the number of interviews to be conducted.

The most tenable hypothesis to explain these findings is that interviewers lost interest and enthusiasm for the work. It may also be that interviewers had some feeling that an aim of the study was to check on their performance, although much reassurance was given that this was not the case and they gave no indication of such concern in interviews conducted with them after the completion of their interviewing. The data indicate that, whatever the motivation, interviewers behaved differently in the earlier interviews than in later ones.

Data from other validity studies of hospitalizations and physician visits were then analyzed to see whether the pattern was replicated and to gain greater understanding of the finding. Data were available from another study on underreporting of visits to physicians during the 2 weeks preceding the interview. A sample was

[4]The reader is cautioned not to interpret the figures in this paper as a measure of net reporting bias, since only underreporting is included. An estimate of the rate of overreporting is not possible with this sample, as it requires a different research design.

Table 31. Number of recorded hospital discharges, median number of reported discharges per interviewer, and rate of underreporting, by number of interviews per interviewer: Survey Research Center

Rank order of interviewers by numbers of interviews taken (lowest to highest)	Recorded discharges	Median number of discharges reported per interviewer	Rate of underreporting discharges per interviewer	
			Median	Mean
1-5	54	4	0	[1]1.4
6-10	237	47	10	10.5
11-15	330	65	12	12.4
16-20	506	100	15	15.6
21-25	697	137	12	[1]12.5

[1]Difference between first five and last five interviewers significant at $p < .05$.

Source: reference 17.

drawn from the records of a large subscription medical care plan in the Detroit area. A systematic sample of visits to clinic physicians was drawn weekly for 8 weeks. With each week comprising a random sample of those visiting the clinic during that particular week, a total of 275 interviews were conducted. Since many respondents had multiple visits, these interviews accounted for a total of 403 visits for the 2 weeks preceding the week of interview.

Ten interviewers were hired by the Survey Research Center for this study. All were comparatively inexperienced. Some had worked briefly on a methodological study; others had worked for a short time on the U.S. decennial census, but most had no previous interviewing experience.

Special training manuals and material were prepared and a supervisor with several years of interviewer training experience conducted the training, assisted by two other experienced field supervisors. Three weeks of training were completed before the actual interviewing started. Classroom training and field assignments were conducted during the first week, and during the second week each interviewer was observed as she conducted interviews at nonsample addresses. The third week consisted of interviewing assignments which interviewers thought were part of the regular study, but which were actually addresses from the telephone directory. During the fieldwork, questionnaires were reviewed in the office and errors were discussed with interviewers.

The questionnaire was nearly identical to that used in the Health Interview Survey. The questions about visits to physicians were as follows: "Last week or the week before, did anyone in the family talk to a doctor or go to a doctor's office or clinic?" Three probe questions were added for specific types of visits which respondents might consider to be outside the scope of the question: (a) "At the time of this visit was the doctor asked for any medical advice for any other member of the family?" (b) "Did anyone in the family get any medical advice from a doctor over the telephone last week or the week before?" (c) "Did anyone in the family see a nurse or technician for shots, X-rays, or other treatment last week or the week before?"

In this study, although interviewers were not given random assignments, the total sample for each week was an independent random sample of visits. It is, therefore, possible to make comparisons of underreporting rates for all interviewers for each of the 5 weeks in which interviewing took place (see table 32).

Table 32 shows a significant increase in underreporting over the 5-week period of the study. The difference between week one and week five is significant at the 5-percent level. There is a decrease in validity of reporting over time. This finding tends to confirm the results shown in table 31. In the first study reporting got progressively worse as the number of interviews taken by an interviewer increased, and in this study underreporting increased as time progressed.

Table 32. Number of recorded physician visits and percent not reported in interviews, by week of interview: Survey Research Center

Interview week	Recorded visits	Percent not reported
1	78	18
2	102	21
3	95	23
4	73	27
5	55	29

Source: reference 17.

However, even though underreporting increased as time progressed, the rate of underreporting was actually lower in the second study when a comparison was made of the rates of underreporting by the number of interviews taken by each interviewer (see table 33). This finding contradicts the conclusions drawn from the data of tables 31 and 32. An explanation may be found in the reasons why interviewers conducted more or fewer interviews in the two studies and in the way the assignments were carried out. In the hospitalization study, interviewers were given weekly assignments by the office and had little to say about the number given because each interviewer worked only in one geographic area and had to take all interviews in that area. Therefore, some had a heavier

Table 33. Number of recorded physician visits and percent not reported in interviews, by number of weekly interviews conducted per interviewer: Survey Research Center

Individual number of interviews per week	Recorded visits	Percent not reported
1	26	31
2	26	38
3	28	29
4	35	20
5	39	26
6	43	26
7	44	26
8	49	20
9	49	20
10	64	14

Source: reference 17.

weekly load than others, regardless of their wishes. In the physician visits study, all interviews were taken in a single area, and each interviewer could take interviews in any part of the area with little increase in cost. Within limits, interviewers were permitted to choose the number of interviews they wished to take each week. Thus, the choice rested with the interviewers. Their choices may reflect a greater involvement or interest in participating in the study, or, since interviewers were paid on an hourly basis, it may reflect a desire to earn more money. Thus, the difference in underreporting rates between interviewers in the two studies may reflect a difference in their motivation. Since interviewers were generally consistent in their choice of a large or small number of interviews each week, the original finding of an increase in underreporting is understandable again in terms of interviewer interest and enthusiasm. Those with low motivation for the job lost interest early, and even the enthusiasm and reporting accuracy of more highly motivated interviewers waned over time in the course of the fieldwork.

Another validity study of hospitalization reports by Marquis and Cannell[50] utilized three different field procedures and thus permits comparison of interviewer performance. For this study, which differs in several respects from the hospitalization study reported earlier, a sample of discharges was selected from 18 hospitals in the Detroit metropolitan area. The design consisted of four orthogonal randomized latin squares. The major sources of variance were five interviewing weeks; five regions of the city; and three field procedures, one control and two experimental. This design provides a base for comparison of interviewer performance that is less confounded with other variables than that of previous studies.

The three field procedures were as follows:

Procedure A.—This was applied to the control group. This procedure involved essentially the same standard Health Interview Survey questionnaire that was used in the other two studies.

Procedure B.—The questions and procedures were the same as in procedure A, except for hospitalizations. The reference period for questions on hospitalizations was a year and a half instead of a year. Probe questions were added and special introductory statements were included.[h]

Procedure C.—The interview questionnaire was identical to that used in procedure A except that no questions on hospitalizations were asked during the interview. A form to be filled out by the family was left with the respondent. Nonresponses were followed up by mail and telephone.

Each interviewer was assigned to two procedures. One group used procedures A and C; the other group used procedures B and C. Twenty interviewers were employed (most of whom had very limited interviewing experience). Inexperienced interviewers were used so that they would not know that the various procedures were different from those customarily used in the Health Interview Survey. The training was conducted by experienced trainers using standard interview methods. There was one full week of training and practice interviews followed by field observation of each interviewer. A comprehensive manual was prepared that specified all techniques. The

[h]This procedure also utilized a mail followup questionnaire designed to pick up further hospitalization reports. The data presented here do not include results of the followup and consist only of reports given in the interim.

training was conducted in two groups and by two trainers, one for those using procedures A and C and the other for those using procedures B and C.

Table 34 shows a pattern percentage of underreporting of hospitalizations for each of the 5 weeks for each of the three procedures. Procedure A (control group) shows a pattern of underreporting much like that found in the study of physician visits. Reporting was poorer as the fieldwork progressed. Procedure B (experimental interviewer) showed a similar pattern with small differences. It may be that the effect of the experimental procedure was to diminish or eliminate the effects of time on interviewer performance. Two factors, the additional probes and the supplementary statements to respondents about the study, may account for this effect. Procedure C (the self-administered form) does not show this pattern and would not be expected to since the interviewer merely handed the hospitalization form to the respondent, asking that it be completed and mailed in. Because the design called for approximately equal numbers of interviews per interviewer, it is not possible to compare reporting rate by number of interviews conducted.

Table 34. Percent of recorded hospitalizations not reported in interviews when procedures A, B, and C were used, by week of interview: Survey Research Center

Week of the interview	Hospitalization interviewing procedure		
	A	B	C
	Percent of hospitalizations not reported		
1	13.7	8.3	14.4
2	11.0	8.6	16.0
3	16.8	9.2	21.2
4	22.1	8.7	10.5
5	23.7	10.0	16.1

Source: reference 18.

There is another bit of evidence in this study which supports a motivation hypothesis. Each interviewer used two procedures: procedure C, which involved the self-administered report of hospitalizations left with respondents to complete and mail, and either procedure A or B, both of which obtained a report of hospitalizations in the interview. The average underreporting rate obtained for each interview can be compared with the average underreporting rate based on the mailed return. The product moment correlation for the reporting rate by interviewer for the control procedure and mailed response procedure is .65; that between the experimental procedure and mailed procedure is .56.

These relationships are surprising, particularly because one reason for using a self-administered procedure was to avoid the interviewer's influence. The relationship is, of course, based on the performance of only 10 interviewers in each group. One interviewer who was singularly unsuccessful in obtaining reports of hospitalizations in either procedure contributed disproportionately to one of the correlations. However, these data indicate, as do previous results, that interviewer behavior varies and that this behavior affects responses. Interviewers who were more successful in stimulating respondents to report hospitalizations during the interview were also more successful in stimulating respondents to better performance in filling out a self-administered form and mailing it to the office.

Another interesting behavior pattern was discovered. The questionnaire for the study of physician visits contained one main question and three followup probe questions designed to make sure that the respondent's concept of physician visits was the same as the interviewer's and to stimulate recall of easily forgotten events. An overall 7-percentage-point improvement in reporting was achieved through use of these followup probe questions, but the data demonstrate that some interviewers used the probe questions quite differently than did others. Not only are there large differences in the amount by which the probe questions improved reporting for different interviewers, but meaningful patterns are present.

The rates of underreporting for the 10 interviewers according to the total number of interviews conducted are shown in table 35. The improvement in reporting through use of the probe questions is shown in the last column of

Table 35. Number of recorded physician visits reported in interviews, percent not reported when procedures A and B were used, and percent of improvement in reporting rate, by individual interviewer

Interviewer number	Number of recorded physician visits reported	Percent not reported		Percent improvement in reporting rate (A - B)
		Main question only (A)	Main question plus probes (B)	
1	26	69	31	39
2	26	38	38	0
3	28	29	29	0
4	35	20	20	0
5	39	26	26	0
6	43	28	26	2
7	44	34	23	11
8	49	33	20	13
9	49	22	20	2
10	64	20	14	6

NOTE: Data previously unpublished.

this table. When only the main question was used there was not a clear relationship between the number of interviews taken and the underreporting rate (rank order correlation of .52). When the probe questions were used there was more of a tendency for interviewers taking an increased number of interviews to have lower reporting rates (rank order correlation of .83). The most interesting finding was related to the effect of probe questions. Only one of the five interviewers recording fewer than 40 sampled visits improved the reporting by using probe questions, but all of those who recorded over 40 visits showed improvement in reporting when the probes were used. The median rate of underreporting is within 1-percentage point for the main question when the top five interviewers are compared with the bottom five interviewers. When all the probe questions were used, there was no change in reporting for the first five interviewers, and a 6-percentage-point improvement in reporting for the five interviewers having the greatest number of reported visits.

Two conclusions are suggested by these data. The first is that interviewers differ in the way they use questions and probes. Some apparently make little use of the probe questions, either failing to ask them or asking them in an incidental manner. For four interviewers the probes failed to elicit any additional information. In contrast, interviewer number 1, who seemed to have placed most of her reliance on probe questions, experienced a 39-percent improvement in reporting when she used the probes. The training and supervision failed to obtain uniformity in the use of the probe questions.

The second conclusion is that, except for the unusual behavior of interviewer number 1, the interviewers who had the largest number of sampled visits (who may have been more interested and motivated) generally made better use of the probe questions. A major interviewer difference in this study and the hospitalization study is the amount of experience in interviewing and in using the Health Interview Survey questionnaire. For the experienced interviewers there was a drop over time in the reporting of known events, suggesting that additional experience did not make for improvement in skill, since skill level was already high. The drop can be accounted for in motivational terms; the novelty of the studies wore off. In the physician visit study, the interviewers were inexperienced. Those who conducted a large number of interviews improved, suggesting that they did profit from longer experience and did gain in skill. This finding plus the earlier motivation hypothesis suggest that interviewers who are motivated improve their performance with added experience. For the less highly motivated interviewers this relationship is not as strong. This interpre-

tation, based only on speculation, would help explain the conflicting results in the time studies.

One final study is relevant. In this study,[46] which did not have validity measures, the dependent variable was the number of health conditions and behaviors reported. Since previous studies suggested that the major problem in reporting health events is underreporting rather than overreporting, the working hypothesis in this study was that high reports are likely to be more accurate than low ones. The study had two experimental interviewing procedures and a control group. The control procedure was fairly similar to the questionnaire used in the Health Interview Survey. Only the control group is used for this analysis. Because the experimental procedures required interviewers to follow rigid rules of interviewing techniques, it was necessary to institute special field supervisory procedures. Each interviewer was observed during her interviewing every week during most of the fieldwork. Attention was focused on interviewing techniques and each interview was discussed with the interviewer.

Average reports for the number of conditions, symptoms, and physician visits respondents reported for themselves and for other family members are shown in table 36. In contrast to data of other studies, these data show higher reporting of events in the second half of the fieldwork, for all but one item. Self-reports for chronic and acute conditions and symptoms were significantly higher during the last 3 weeks of the study than during the first 3 weeks.

Again a motivational hypothesis explains these findings. In this study, in contrast to the preceding ones, attention was paid to the interviewer's performance in interviewing. In previous studies interviewers were rated on the quality of the completed questionnaire, while for this study the reward was for good interviewing performance. It seems likely that both different criteria and greater personal attention led to greater skill and increased motivation in performing the interviewing task.

Table 36. Average frequency of health reports during first 3 weeks and second 3 weeks of the interviewing period using procedures A and B, and improvement in reporting during second 3-week period, by respondent status and health event

Respondent status and health event	Average events reported		
	First 3 weeks (A)	Second 3 weeks (B)	Improvement during second 3 weeks (B – A)
Respondent reporting for self			
Chronic and acute conditions	1.87	2.53	[1].66
Symptoms	4.88	5.83	[1].95
Physician visits	0.34	0.60	.26
Respondent reporting for other family members			
Chronic and acute conditions	1.50	1.26	–.24
Physician visits	0.14	0.26	.12

[1] $p < .01$.

CONCLUSIONS AND DISCUSSION

The evidence from these data indicates that the interviewers generally did not improve their performance as they gained experience, at least not during a single study. In fact, performance began to deteriorate almost immediately after training and, at least in some cases, dropped significantly within a few weeks. Contrary to expectations, this deterioration occurred among both experienced and inexperienced interviewers.

The fragmentary data are susceptible to various interpretations. One explanation is that reporting bias is related to the interviewer's inability or lack of incentive to encourage enthusiastic performance by the respondent and that, over time, interviewers become less conscientious in their use of interviewing techniques. Because performance dropped as the interviewing period progressed, it would seem that performance is related more to motivational level than to ability to perform the task. Even though the nature of the behavioral change in the interviewer is not known, there is evidence that the manner in which the questions are asked may change over time. Whatever the mechanisms are, the interviewer's behavior apparently has an effect on the respondent's

behavior, both in the reporting during the interview itself and in the respondent's subsequent performance in filling out and mailing a self-administered form.

It is, of course, no surprise that lack of incentive or interest in performing the interviewing task results in poor performance, but it is surprising that the interviewer's performance dropped as rapidly as these findings indicate.

It appears that interviewers were being reinforced for part of their role performance, and that the reinforcement brought about improvement. However, adequate performance in stimulating respondent reporting behavior and in proper use of interviewing techniques was not reinforced and, accordingly, deteriorated. Performance in interviewing was not reinforced because observation of interviewers was not continued after training, and good or poor performance could not be identified by a review of the completed questionnaire. (It is instructive to note that in one of the studies reported here the interviewer who was among the poorest in obtaining known hospitalizations and whose underreporting rate rose sharply over the interviewing period was selected as the best interviewer and was promoted to an office supervisory job because of her "excellent" work.)

There are two major implications in these findings. The first is that research is needed to identify explicitly those interviewer behaviors that relate to adequate reporting. Studies that concentrate on interviewer-respondent interaction, especially through analysis of recorded interviews, can provide information for better understanding of some of these factors. A second implication is the need for training and supervisory methods that will reinforce adequate interviewer performance in the application of interviewing techniques.

Studies frequently show a relationship between the information reported in interviews and the attitude of, or expectations of, the interviewer. Interviewer bias studies abound, but all relate characteristics of the interviewer (background characteristics, attitudes, perceptions, and expectations) to the reporting of respondents. The data presented here point to the importance of the interviewer's continued motivation to do a conscientious job of interviewing. The suggestion is that as fieldwork progresses, interviewers become less careful or conscientious in using the techniques they were trained to use. Furthermore, there is evidence that interviewers who performed well in the interview also inspired their respondents to perform well, as was shown in the adequacy of respondent reporting of hospitalizations in a self-administrative procedure. Good interaction improves not only the technical aspects of the interview, but also stimulates a high level of respondent role performance which extends in time beyond the interaction itself.

Interviewer training needs to include devices for building the interviewer's enthusiasm for the job as well as procedures for training in the proper use of techniques. Attention to performance cannot stop when interviewer training is completed. Training needs to be evaluated and good performance reinforced continually during the period of production interviewing.

THE USE OF VERBAL REINFORCEMENT IN INTERVIEWS AND ITS DATA ACCURACY

The purpose of this section is to describe interviewer verbal reinforcement as it occurs in interviews, to conceptualize it in the same manner as other major interviewer behaviors, and to look at some of the relevant research studies that provide information about how verbal reinforcement influences the accuracy of interview data.

The major category of interviewer behavior of concern here is interviewer feedback. Interviewer feedback consists of evaluation statements that the interviewer makes after the respondent says something. The subject of feedback may be approached through two basic questions: "How can interviewer feedback be used to improve data?" and "How can it be arranged so as not to introduce unwanted bias?" The following discussion will be limited to interviewer verbal reinforcing statements which represent only one kind of interviewer feedback.

Probes and statements that are intended to build rapport are not treated directly. While this restriction is undesirable, it is necessitated by the fact that research conducted at the Survey Research Center and elsewhere has dealt mainly with task-oriented verbal reinforcing statements.

RESEARCH ON INTERVIEWER REINFORCEMENT

Since Pavlov's famous experiments, there have been numerous studies of the effects of reward and reinforcement on learning and performance. Much of this research has been carried out with animals and occasionally with children. It has been assumed that the principles of reinforcement derived from these laboratory studies could be projected to adult human beings. This assumption has been made explicitly by many writers, most notably Dollard and Miller[51] and Skinner.[52]

Operant Conditioning Studies With Verbal Reinforcement

Outside the psychology laboratory two important kinds of practices, programed instruction (or teaching machines) and behavior modification therapy, demonstrate the powerful effects of properly scheduled reinforcement on human performance. Some practitioners in the fields of education and psychotherapy have felt that traditional methods of producing learning or behavior change (e.g., lecture, introspection) are inefficient. They have met some success with new techniques based on tightly controlled operant conditioning techniques where reinforcement is contingent on the respondent's actual behavior.

These studies point to the importance of feedback on human performance under a variety of conditions. However, the experimental situations mentioned above (sentence construction, free association, and casual conversation) are not like a typical survey interview. Several studies have been conducted in interview settings. The first set of studies shows that interviewer reinforcement has great effects on the amount of respondent behavior. The second group of studies shows that interviewer reinforcement can produce an interviewer bias effect. Finally, the study by Marquis, Cannell, and Laurent[53] indicates that, under certain conditions, interviewer feedback can be used to increase the accuracy of information given by respondents.

Effects of Feedback on Amount Reported

A study by Marquis and Cannell[50] shows that the presence or absence of interviewer verbal feedback has a significant effect on the amount of health information reported by a respondent in a household interview about family health. In this study a probability sample of moderate-income females between the ages of 17 and 65 and living within the city limits of Detroit was interviewed. They were asked questions about their own health, about their use of different kinds of medical services, and about the health of another person in the household. Although three experimental groups were used in this study, two groups of respondents are of interest here: (a) the group of respondents who were reinforced every time they reported a symptom or health condition either for themselves or for the other person in the household for whom they were reporting; and (b) the group of respondents who were interviewed using the same questions but who received no reinforcement or feedback from the interviewer.[i]

For the first experimental group the kinds of reinforcing statements which the interviewer used are outlined here. Words in parentheses could be used, omitted, or interchanged by the interviewer:

(Yes.)	That's the kind of information we need.	
(Um-hmm.)	We're interested in that.	
(I see.)	You have	(name of condition)

There were some other small differences in the interviewing procedures for the two groups and these are described in detail elsewhere.[50]

[i]The third group was also interviewed without reinforcement. A list of symptoms, which appeared at the beginning of the other two kinds of interviews, was inserted near the end of the third treatment interview. The purpose of this procedure was to test the sensitizing effects of the symptoms list on later reporting. No main effect was found on the amount reported later in the interview, although it may be necessary to begin a reinforcement interview with a symptoms list which allows all respondents to report some sort of sickness and to receive reinforcement for their reporting.

The main dependent variable in the study was the number of chronic and acute conditions which the respondent reported for herself. Following the logic outlined by Cannell, Fowler, and Marquis[46] (appendix II), it was assumed that the more chronic and acute conditions reported, the more valid the overall data. The group of respondents receiving reinforcement reported an average of 2.74 conditions per person. The nonreinforced group reported an average of 2.20 conditions. The data indicate that the reinforced group of respondents reported about 25 percent more chronic and acute conditions than did the nonreinforced group. The difference is significant at the .02 level of confidence. The same magnitude of effect was obtained for reporting chronic and acute conditions for the other person in the household. The number of conditions reported by proxy averaged 1.88 in the experimental interview and 1.43 in the control interview. The experimental or reinforcement interview obtained about 24 percent more chronic and acute conditions reported by the proxy respondent and the difference is significant at the .01 level of confidence. Other results indicated that the reinforcement effect was general rather than limited to one kind of condition. For example, the reinforced respondents reported about 25 percent more medically attended conditions for themselves and about 55 percent more of the highly embarrassing symptoms than did the nonreinforced group. The group that received feedback from the interviewer did not report more visits to physicians than did the nonreinforced group. This finding is discussed in more detail in the original report by Marquis and Cannell.[50]

Kanfer and McBrearty[32] interviewed 32 female undergraduates about 4 predetermined topics. During the first part of the interview the women were handed cards listing each of the topics and were asked to talk about each one. The experimenter sat in the room but kept completely silent. During the second part of the session the experimenter reinforced the respondent whenever she talked about a predetermined two of the four possible topics. Reinforcement consisted of a "posture of interest" including smiles and the phrases "I see," "Um-hmm," and "Yes." When the women talked about the remaining two topics, the experimenter said and did nothing. The subjects spent more time talking about the topics that were reinforced than about those topics that were not reinforced.

These two studies indicate that interviewer feedback can have a great effect on how much is said during the questioning. It should be noted that both of these studies involved maximum contrasts in interviewer feedback; that is, in one group or at one time interviewers used verbal reinforcement and for another group or at another time interviewer feedback was completely nonexistent. While these studies show that interviewer feedback can have important effects on the amount reported, they indicate little about the nature of the effects or what kinds of feedback procedures are most effective.

Use of Feedback To Increase Accuracy

The study to be discussed here is one in which the data indicate that interviewer feedback procedures can have beneficial effects on survey data accuracy. It should be remembered that this is a single study dealing with fact responses, and that the results are tentative and possibly of limited generalization. This study by Marquis, Cannell, and Laurent[53] was carried out under contract with The National Center for Health Statistics. In an unpublished study by Marquis and Laurent,[54] a number of interviewing variables were tested but the discussion that follows is restricted to the effects of reinforcement on initial interviews and reinterviews using short (nonredundant) questions. The sample respondents were white females between the ages of 17 and 65, who were residents of the greater metropolitan Detroit area. The respondents were selected on a weighted basis from persons who had come into a prepaid health clinic during a 6-month period prior to interviewing. During the time each patient was in the clinic a physician filled out a checklist indicating which of 13 chronic conditions the patient definitely or probably had, and about which of the checklist conditions the physician had no information. The physician obtained information about the chronic condition from the patient's record and from his own knowledge of the patient's medical condition.

He did not interview the patient directly about the specific conditions listed on the form.

The dependent variable used in the study was the accuracy with which respondents, in a series of subsequent household interviews, reported having had or not having had each of the 13 aforementioned chronic conditions. Two types of respondent error were (a) underreporting, in which case the respondent failed to mention a condition which the doctor said she had; and (b) overreporting, where the respondent mentioned a condition which the doctor said she did not have.

It was not expected that the respondent and physician would agree entirely on the state of the respondent's health. However, what was expected, for the reporting of these 13 very common chronic conditions, was that an experimental interviewing procedure that obtained low average rates of respondent-physician disagreement obtained more valid data than an interview procedure yielding high rates of disagreement. The study included a reinforced group of respondents and a nonreinforced group. All respondents received the same kind of questionnaire composed of straightforward short questions. The verbal reinforcement used in the reinforced group was similar to that described in connection with the study by Marquis and Cannell.[50] Half of the respondents were reinterviewed approximately 2 weeks after the initial interview. Respondents in the experimental group received reinforcement for reporting all kinds of health events, not just symptoms and conditions as in the 1971 study just cited. The results obtained in the analysis of the 1972 study data were not those initially anticipated.[53] They indicated that the effects of different ways of conducting interviews were mediated by the level of education of the respondent. Procedures that increased accuracy and reduced bias in the low education group (had not completed the 12th grade) had opposite effects in the high education group. The reader should consult the full report of this study[j] for a more detailed treatment of all data and all experimental techniques.

The discussion of the results of this study is limited to four groups of respondents, those reinforced and those not reinforced in each of the two education groups, all of whom were interviewed with short questions.[k]

The underreporting index scores shown in table 37 indicate that, for respondents who had not finished high school, the use of reinforcement in interviews significantly reduced the underreporting of chronic conditions in the original interview. However, reinforcement significantly increased underreporting error for the more educated group. The drop from .387 to .225 in underreporting brought about by reinforcement in the less educated group represents a decrease of about 42 percent; the percentage increase of underreporting in the higher education group due to reinforcement (from .443 to .575) was approximately 30 percent. Both of these differences are significant at the 5-percent level.

The overreporting errors were generally low (7 to 14 percent) and were less affected by reinforcement. The slight trends in the data suggest that reinforcement reduced overreporting in both education groups. One hypothesis in this study was that reinforcement by the interviewer each time the respondent reported an illness would result in the respondent "inventing" sicknesses in order to gain further approval of the interviewer. Since the reinforcement interview produced fewer false positive responses than the interview that did not include reinforcement, the trend was counter to that expected.

The total error index scores shown in table 37 represent the unweighted sum of the individual underreporting and overreporting scores. Since underreporting scores had a higher mean and variance than overreporting scores, the total error score reflects the effects of underreporting due to reinforcement and education level. For the less educated group, reinforcement signifi-

[j] Copies of this study are available from the National Center for Health Statistics, *Vital and Health Statistics*, PHS Pub. No. 1000-Series 2-No. 45.

[k] The reinforcement contrast was also tested using a questionnaire which contained a mixture of long and short questions. For reasons not discussed here, the effect of reinforcement with long questions was ambiguous. It is clear, however, that the combination of reinforcement with long questions does not result in marked improvement in data accuracy.

Table 37. Error rates for chronic condition reporting in original interviews and reinterviews, by use of reinforcement and education level of respondent

Respondent's education level	Original interview		Reinterview	
	Reinforced	Not reinforced	Reinforced	Not reinforced
Low education level				
Number of respondents	50	41	25	20
Total error	.338	.530	.361	.566
Underreporting	.225	.387	.230	.400
Overreporting	.113	.143	.131	.166
High education level				
Number of respondents	38	53	20	27
Total error	.644	.534	.603	.615
Underreporting	.575	.443	.521	.495
Overreporting	.069	.091	.082	.120

cantly lowered the total error; for the more educated group, reinforcement significantly increased the overall error.

The trends in the reinterview data were in general similar to those observed in the data as collected originally. The questions were approximately the same as those asked the first time. Respondents who were in the reinforced group originally were also reinforced during the reinterview. Similarly, respondents initially in the nonreinforced group, did not receive reinforcement the second time. Thus, the reinterview data reflect the cumulative effects of either reinforcement or nonreinforcement.

While the results of this study are not as clear as one might wish, they do indicate that certain kinds of reinforcement procedures can have a marked effect on the accuracy of the estimates made for a population from survey data. It would appear that less educated respondents rely on interviewer cues to direct their reporting performance. Thus, appropriate use of reinforcement, probes, and other feedback by the interviewer can aid recall and reporting accuracy. Perhaps because of the perceived social status discrepancy, feedback is accepted as appropriate and actually welcomed. Performance is generally good in an appropriately conducted initial interview, but there is little additional benefit from a reinterview.

More educated respondents carry out the reporting task according to their own understanding of it rather than relying on cues from the interviewer. Reinforcement under such circumstances may be perceived by the respondent as inappropriate, unnecessary, or even condescending. The more educated respondent appears to have a tendency to underreport chronic conditions, possibly due to a stronger conservatism or social desirability bias. While these findings are somewhat speculative, they do stress the importance of such "human" characteristics as memory, recall, cognition, social status, and intellectual ability that are often overlooked in methodological studies.

The studies cited to this point indicate that:

a. Interviewer feedback makes a difference in respondent performance.
b. Interviewer feedback can bias respondent answers.
c. Interviewer feedback can increase response validity.

However, these findings are based on experimental interview studies with unorthodox ways

of programing feedback and do not indicate how feedback is used in the normal household interview.

Use of Feedback in a Nonexperimental Interview

Marquis and Cannell[2] obtained tape recordings of 181 interviews with employed male respondents about employment-related topics. The respondents included both black and white persons ranging in age from 18 to 64. The four interviewers were white middle-class female residents of suburban Detroit.

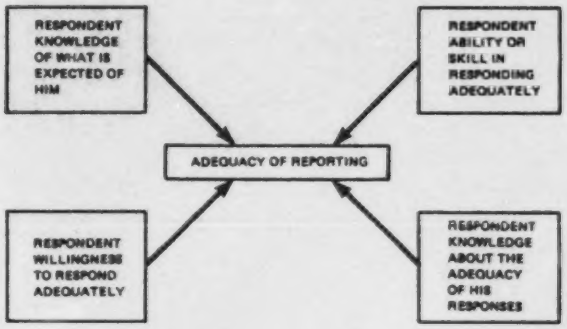
The tape recordings of each interview were analyzed by identifying each instance of verbal behavior on the part of the interviewer or the respondent and assigning it 1 of 36 possible behavior codes. Thus, each interview was transformed into a series of code symbols which represented the kinds of behavior in which the interviewer and respondent engaged. The data indicated that 23 percent of all coded interviewer behavior consisted of feedback. Feedback, as a category of interviewer behavior, occurred as frequently as probing. The only kind of interviewer behavior that occurred more frequently than feedback was question-asking (this accounted for about 37 percent of all interviewer behavior). These data indicate that interviewer feedback is, indeed, a large part of interviewer behavior.[1]

Data presented in table 29 are the basis of discussion of the use of reinforcement by interviewers. The table shows a very surprising finding: The probability of feedback after an inadequate answer was almost the same as the probability of feedback after an adequate answer. Even more surprising is that interviewers on the average reinforced over half of the respondents' refusals to answer. This kind of pattern discloses that interviewer feedback may be used in an inappropriate manner. These data indicate that interviewers not only gave positive evaluation responses when the respondent replied adequately, but they also said positive things in situations where they may have been under some tension or may have felt the need to build rapport in the interview. The rapport hypothesis about the use of feedback is supported by the finding that feedback was given after refusals to answer and after answers that did not meet the objectives of the question.

The authors have never systematically discussed these data with interviewers, but were this to be done, it might be expected that interviewers would insist that the difficulty of their job makes it necessary for them to build rapport with respondents by using positive reinforcing statements when tension is felt during the interview. The authors would probably reply that at least one experimental study has shown that validity can be improved by reinforcing only adequate answers and that this pattern should be followed. Possibly these two divergent hypotheses about the effective use of interviewer reinforcing statements can be tested experimentally. At this point there is only limited empirical support for the authors' position and intuitive, common-sense support for the interviewers' position.

Discussion

The data indicate that the survey researcher should be concerned about the feedback styles used by interviewers. The remainder of this section discusses ideas and research concerned with reinforcement effects in the personal interview. Before going further into the hypothesized effects of verbal reinforcement on respondent performance, a schematic diagram of variables that may influence reporting quality might be helpful:


RESPONDENT KNOWLEDGE OF WHAT IS EXPECTED OF HIM
RESPONDENT ABILITY OR SKILL IN RESPONDING ADEQUATELY
ADEQUACY OF REPORTING
RESPONDENT WILLINGNESS TO RESPOND ADEQUATELY
RESPONDENT KNOWLEDGE ABOUT THE ADEQUACY OF HIS RESPONSES

[1] As this report goes to press, results from a small pilot observation study (N=23), conducted on a different kind of interview (urban problems) and with more experienced interviewers, indicate that the proportion of interviewer behavior devoted to short, positive feedback statements was 10 to 15 percent. That this percentage range is considerably lower than the proportion obtained in the labor force study suggests the need for further research.

The diagram implies that there are four hypothetical personal characteristics (other than having correct answers available in memory) that will affect the accuracy and completeness of reporting. These consist of two knowledge (cognitive) variables—knowledge of what is expected and knowledge of how well one's responses are meeting these expectations—a skill variable, and a motivational variable.

This diagram implies that if the respondent knows what is expected of him, has the ability to do what is expected, can tell how adequate his responses are vis à vis the task requirements, and wants to do well, then the data he gives will be accurate and complete. The purpose of the discussion that follows is to explore how verbal reinforcement procedures and alternative procedures affect each of these hypothetical variables.

The tentative conclusion reached is that positive verbal reinforcement contingent on adequate answers provides a wide variety of desirable effects on these intervening constructs, while other procedures that might be used tend to have more limited effects. The effects of verbal reinforcement at any particular stage are not totally clear and derivations from theory suggest reinforcement may also have some counterproductive effects in certain situations.

THREE KINDS OF INTERVIEWER VERBAL REINFORCEMENT EFFECTS

The effects of verbal reinforcement on respondents can be divided into three categories: a cognitive effect, a conditioning effect, and a motivational effect. There can be a great deal of overlap among these categories but, for ease of presentation, the three kinds of effects will be treated as if they are independent.

Interviewer verbal reinforcement has a cognitive effect on respondents because it supplies information about the expectations of the interviewer and how adequately the respondent is meeting these expectations by his answering behavior. For example, if the interviewer asks whether the respondent has ever had a headache, the respondent says "Yes," and the interviewer uses a reinforcing statement such as, "Good, that's the kind of information we need." The interviewer has indirectly told the respondent that she expects him to report his minor illnesses and that the response he just gave was a good one in terms of meeting the objectives of the health interview. This information-giving aspect of verbal reinforcing statements which changes (or maintains) the respondent's intellectual understanding of the respondent role requirements and the adequacy of his responses is described as a cognitive effect.

The second effect that interviewer feedback may have is referred to here as "instrumental conditioning." This effect is important in many studies of the psychology of learning. In the simplified model of the interview just described, the interviewer's evaluation comes immediately after the respondent has given a response. This sort of pairing of an evaluation with a response can have reinforcing properties, that is, it can alter the frequency of the behavior that immediately preceded it. Thus, the evaluation process has the potential of strengthening or weakening the probability of eliciting that behavior or similar behavior on subsequent trials. Also, through the process of differential reinforcement of successive approximations, the reinforcement procedure can establish a new kind of response class or behavior in which the respondent would not normally engage on his own.

The third possible effect of interviewer feedback is motivational. In this case feedback affects the intensity or psychological effort which the respondent allocates to his reporting task and to other behavior that may interfere with the adequate performance of his respondent task.

Cognitive Effects of Reinforcement

Two kinds of knowing or understanding are thought to influence respondent performance: understanding what is expected (e.g., knowledge of the proper respondent role) and knowing when answers meet and do not meet those expectations. Two ideas are hypothesized: (a) such knowledge is sometimes very helpful but is neither necessary nor sufficient for good performance in some conditions; and (b) reinforcement plays an important role in helping the respondent to acquire both kinds of knowledge and is a more effective procedure when compared to more conventional teaching techniques.

Before exploring the effects of reinforcement on knowledge, it should be noted that many people maintain that the most effective way to increase a person's knowledge and understanding is to teach him by direct (e.g., to lecture) rather than by indirect (e.g., using feedback) methods. Direct approaches to teaching which are not dependent on respondent performance are not uncommon in survey interview settings. For example, before an interviewer arrives at the doorstep, the potential respondent has often received a letter or a brochure that explains the survey and describes what is wanted from the respondent. Often at the beginning of the interview, an explanation of the goals of the research is given, accompanied by specific appeals for accuracy, completeness, or candor.

Very little is known about the effects of this common, direct approach to teaching a respondent. However, if it were possible to extrapolate from the lecture analogy and to research these effects, it might be found that often the "students" had not learned or understood or would not be able to verbalize what they had been expected to learn. Two studies obtained some relevant data. In a study by Cannell, Fowler, and Marquis[55] respondents were reinterviewed 1 to 3 days after an initial health interview. Two questions were asked of 412 respondents to find out how well they had understood their role. The answers were distributed as follows:

Q. 26. Did the interviewer want you to be exact in the answers you gave or were general ideas good enough?

Respondent's answer	Percent distribution
"Exact"	55
"Some of each"	5
"General"	35
Not ascertained	5

Q. 27 Did she (interviewer) want everything, no matter how small it was, or was she interested only in fairly important things?

Respondent's answer	Percent distribution
"Everything"	76
"Important things"	19
Not ascertained	5

A similar question was asked of 428 respondents in another study by Marquis and Cannell[50] at the end of the health interview rather than at a followup interview:

Q. 17. Will people think we want them to report all their illnesses or only the important ones?

Respondent's answer	Percent distribution
"Report all illnesses":	
Gave correct reasons[m]	31
Gave incorrect reasons	28
"Report only important ones"	41

The data indicate that despite introductory letters and brochures, interviewer explanations, and the actual experience of the interview, between 20 and 40 percent of respondents in these health studies clearly did not understand the respondent role correctly.

Some data are available to indicate what techniques are effective in changing respondent understanding. In an unpublished Survey Research Center study by Cannell, Fowler, and Marquis,[56] different kinds of brochures and introductions were used. The effect on respondent understanding of an unattractive but informative (control) brochure was compared with that of:

a. A brochure indicating the kinds of questions to be asked, the interviewer's role, the respondent's role, and the importance of accuracy in reporting;
b. A brochure mentioning uses of data and stressing the benefits that might result from the survey;

[m] These reasons indicated that reporting all illness was what the survey, interviewer, or Government wanted or that reporting all illnesses would result in good data.

c. A calendar on which the respondent might write family sickness information for two weeks prior to the health interview; and
d. A set of actual questions to be asked during the interview, asking the respondent to think about them and consult records and family members for accurate information.

Results of pretests were so disappointing that full scale evaluations of the effects of these communication/teaching devices were not undertaken. The experimental materials, whether used singly or in combination, made no difference in respondent knowledge or reporting performance.

It is currently believed that teaching may be more effective if methods of programed instruction are used. In contrast to the brochure or lecture techniques, one essential feature of the method of programed instruction is immediate feedback about the adequacy of each answer the student gives. One characteristic of reinforcement is that it can provide immediate information to the respondent about the adequacy of his answer. The first SRC reinforcement study mentioned above[55] gave positive verbal reinforcement for adequate answers (see report of that study for feedback procedures and definition of adequate answers) to one group and not to another group. Both groups received the same advance letter and explanation of the study by the interviewer. The following data indicate that the reinforced group had a slightly better understanding of the respondent's task than did the group not receiving feedback for adequate answers:

Q. 17. Will people think we want them to report all their illnesses or only the important ones?

Respondent's answer	Reinforced (151 respondents)	Not reinforced (277 respondents)
	Percent distribution	
Total	100	100
"Report all illnesses":		
Gave correct reasons	37	27
Gave incorrect reasons	28	28
"Report only important ones"	35	45

Therefore, a reinforcement procedure does have at least a small potential for teaching a respondent. However, reinforcement effects appear to be stronger in the area of eliciting adequate responses than in teaching what is expected, since the effects of reinforcement on the actual amount reported are greater than the effects on understanding the task requirements.

Two remaining points about verbal reinforcement and knowledge effects on reporting performance need to be discussed: (a) certain types of reinforcement procedures used in the laboratory experiments mentioned above may be less effective than the procedures used in the 1967 study; and (b) the effects of the verbal reinforcement procedures used in the 1967 study did not appear to be mediated by the degree of the respondent's knowledge of expectancies (awareness of respondent task requirements).

There is a lively controversy over whether or not the respondent must know what is expected of him in order to perform well. In other words, is complete awareness of what is expected a necessary, although possibly insufficient, condition for accurate reporting? Despite the finding that reinforcement appeared to increase understanding of the response requirements, the data from the 1967 study mentioned above show no correlation between reporting and awareness as measured by question 17. This is true for the reinforced group as well as for the nonreinforced group. Yet the reinforced group showed superior performance in reporting their health conditions. Cannell, Fowler, and Marquis reported a slightly positive, but not statistically significant, relationship between awareness and good reporting performance.[46] Fowler reanalyzed the latter data and found that the relationship between awareness (measured by one of the two questions) and the amount of information reported was large for respondents with at least a high school education.[44] That awareness, however, did not predict reporting behavior for respondents with less education, nor did data from the second awareness question predict behavior. Possibly the reinforcement effect, if it is cognitively mediated, is produced by letting the respondent know when his responses are adequate rather than by giving him knowledge of general task requirements. This

implies that the respondent need not understand exactly what is expected of him to respond well.

Recent experiments in social psychology have tried to explore the relative effects of awareness of task requirements and reinforcement of performance. A detailed treatment of these findings is beyond the scope of this report, but some of the main issues and findings are presented at the end of this section where further experimentation is discussed. It appears that awareness can be helpful when the respondent (a) can tell when his single responses are meeting task requirements, (b) has the skill to perform what is requested, and (c) has the will (motivation) to perform.

Feedback used in the 1967 studies may have provided two types of information, and achieved the increase in reporting frequencies because of this double-barreled effect. In some of the laboratory experiments mentioned earlier, reinforcing statements contingent on correct answers consisted of "Yes," "Mm-hmm," "O.K.," or "Good." The 1967 studies used verbal reinforcements which contained more information, statements such as: "Mm-hmm, we need to know that," or "I see, that's the kind of information we need." Probably the two types of statement convey different amounts and kinds of information. The simple statements apparently convey information about response adequacy, leaving it to the respondent to infer the interviewer's goals. The longer statements possibly make it easier for the respondent to arrive at some knowledge of interviewer expectations. This hypothesis would be relatively easy to test in the laboratory or in field interviews.

These conceptions of how reinforcement teaches the respondent to understand his proper role point to the desirability of using more efficient techniques which do not rely so heavily on the respondent's ability to figure out what is expected. Further research seems to be needed in order to test whether the respondent's understanding of his role is necessary for good reporting or whether knowledge of the adequacy of his responses is, in itself, sufficient. If knowledge of expectations is found to be important, it might be effective to develop a method that informs the respondent of his expected performance quality and at the same time, through reinforcement procedures, gives him immediate information about the adequacy of his answers.

The effects of reinforcement on two kinds of respondent knowledge and the role of knowledge in producing good survey data have been touched upon above. Other variables can also influence data quality. In the following sections reinforcement is discussed in the context of two other variables, skill level and motivation.

Conditioning Effects of Reinforcement

Experiments cited earlier show that giving reinforcement immediately after some behavior increases the probability that such behavior will recur. This pattern is defined here as an operant conditioning effect, or more simply a conditioning effect. In the tradition of B. F. Skinner, the arrangement of reinforcement contingencies to produce the conditioning effect minimizes consideration of the intervening variables that might help explain the obtained results. Thus, the following paragraphs omit speculation about how interviewer reinforcement may affect cognitions and motivations to achieve a performance effect. They relate to such problems as identifying the correct behavior to reinforce and the circumstances under which a conditioning procedure seems appropriate.

When to use a conditioning procedure.—The following discussion is based on the premise that conditioning effects of reinforcement should be used (a) when respondents do not have a clear understanding of how to perform effectively and cannot be given this understanding by mere explanation or appeals for good performance, (b) when the respondent does not possess the ability or skill for good reporting regardless of whether or not he understands what is wanted, or (c) when the respondent is performing in a manner which is less than optimal (for example, he wants to do something besides answer questions accurately and completely). The implication is that if the respondent understands what he is supposed to do, has the ability to do it, knows when he is doing it, and wants to do it, there is no reason to introduce a conditioning procedure into the interview. However, all of these conditions are seldom met in any personal interview situation.

The general conditioning pattern suggested for personal interviews is the scheduling of reinforcing statements only after adequate answers. In this way the frequency of adequate answers can be expected to increase. For example, an adequate answer appears fairly easy to elicit with increasing frequency when the following instruction is given: "When I ask an open-ended question, please tell me *all* you can think of, using specific examples." This is accomplished by an interviewer using a reinforcing statement every time the respondent gives an acceptable idea or example in response to an open-ended question. The arrangements to produce an increase in valid answers, especially to forced-choice questions, are not that easily conceptualized. Methods that change the reinforcement contingency rules to reward responses that come closer and closer to approximating some goal have been shown to be effective and, theoretically, could be used in the personal interview setting.

Another use of conditioning procedures is to maintain an acceptable level of respondent performance. Once the respondent has been taught appropriate response behavior, it is desirable for him to continue to answer acceptably. According to the operant conditioning literature, a previously reinforced response (especially if it has been reinforced every time it has occurred) will be given less and less frequently if it is no longer reinforced. Thus, another function of a verbal reinforcement procedure is to prevent extinction of appropriate behavior once it has been established. A schedule of reinforcement that uses positive statements after at least some of the desired behavior is probably adequate to serve the response maintenance function. Procedures that omit reinforcing statements or use them in a random way probably cannot maintain any performance pattern that has been established earlier.

Problems of conditioning procedures for teaching.—A conditioning effect of reinforcement involves the respondent learning how to perform or improve his performance on some task. He may or may not be aware of his resulting increase in ability. For a feedback technique to be effective, the interviewer (or study designer) must be aware of what is to be taught or improved, be able to recognize when a respondent performs the desired behavior, and be able to give appropriate verbal reinforcement immediately after the respondent behavior occurs. These are very stringent conditions, and in the usual interview situation they cannot be met. One major obstacle is the difficulty of determining immediately whether a particular respondent behavior is desirable or not. Without knowing which responses to reinforce, a proper conditioning effect cannot be obtained.

The dangers of trying to teach one concept but actually teaching something else can be illustrated by a recent series of experiments. Cannell and Marquis originally diagnosed the problem of error in reporting chronic and acute health conditions as an underreporting or "failure to report" problem. Subsequently, in their 1967 study,[50] they reinforced every instance where the respondent reported a symptom or a chronic or acute health condition, and obtained a 25-percent increase in the number of symptoms and conditions reported. A subsequent look at unpublished data about chronic condition reporting from other sources suggested that overreporting might be more of a problem than underreporting for this kind of health information. Therefore, the conditioning procedure used may have decreased the accuracy of the data in that study because it led to increased overreporting errors. Therefore, Marquis, Cannell, and Laurent[53] carried out a second study using independent records from physicians as indicators of the presence or absence of chronic conditions for particular respondents. Respondents received approximately the same kind of reinforcement as in the earlier study. The results mentioned earlier in connection with the discussion of table 42 did not show the expected biased conditioning effect. If the conditioning effect were the only effect operating, it would be expected that overreporting (reporting of nonexistent chronic conditions) would increase. The data indicated just the opposite. The short-question reinforcement interview had its major effect in reducing overall reporting error by reducing the amount of overreporting. Exactly why this happened is uncertain. Possibly no conditioning effect occurred, and the reduction in error was due to increased knowledge or motivation effects.

Another possibility is that a concept other than a positive or sickness-reporting response set was taught.[n] The main point, however, is the difficulty in predicting what sort of consequences will arise when interviewers use reinforcement in a particular contingency schedule. While the above example shows that reinforcement procedures had a beneficial effect on the validity of health reporting, the effect may not have been for the reason originally hypothesized.

Substitute for conditioning procedures when the problem is one of low respondent ability.—Probably one of the biggest problems in survey interviewing concerns the respondent's ability to recall factual information accurately. The most dramatic example of a memory problem is shown by Cannell and Fowler[1] in the reporting of visits to a physician during the 2 weeks preceding the interview. Using a standard question designed to find out how many times the respondent had consulted a physician during this time, 21 percent of the visits known to have occurred within the last week were not reported and 38 percent of the visits known to have occurred during the second week prior to the interview were not mentioned. Inability to remember increased over 80 percent from the first to the second week preceding the interview.

It has been hypothesized that a conditioning procedure might help the respondent learn to distinguish correct from incorrect memory representations of an event such as a physician visit. In theory, this kind of teaching seems possible but it has not yet been demonstrated as feasible in the interview setting.

A nonreinforcement interviewing approach might be considered when ability to remember correctly presents problems for survey data accuracy. The recommended approach involves:

a. Making the recall task simpler so that memory ability is less important for validity; and
b. Using repeated trials so that an initially faulty recall decision may be corrected.

Evidence for the possible efficacy of this task simplification and repeated trials approach is presented in another section of this report, "Memory and Information Retrieval in the Interview." It should be remembered, however, that if response error is thought to be due to variables other than memory failure, e.g., misunderstanding or fear of embarrassment, a reinforcement procedure might be more appropriate.

Motivational Effects of Reinforcement

The use of feedback statements, *regardless of the rate or the contingency rule for their use,* may have motivational consequences. The motivational effects are probably not reflected directly in the awareness or consciousness of the respondent, and hence may be difficult to detect by questioning.

One motivational consequence has been mentioned previously. In standard interviews it was observed that feedback statements were used often in situations where there may have been tension or negative feelings. Implicit here is the hypothesis that positive statements by the interviewer can reduce the respondent's tension or hostile feelings about the interview. This hypothesis should be tested.

Psychological theory suggests at least three possible motivational effects of positive feedback statements. These statements might:

a. Affect the general level of motivation drive;
b. Strengthen or weaken levels of specific motives which facilitate or inhibit reporting performance; or
c. Affect the degree of approval the respondent has for the interviewer. The effect of approval on performance is ambiguous and is discussed briefly below.

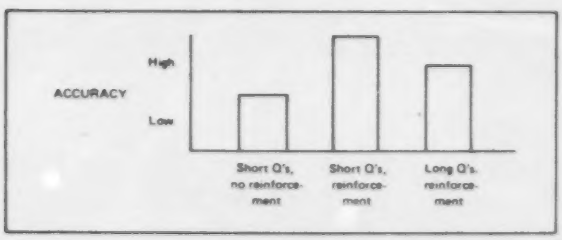
General motivation.—Some theories of motivation include the idea of "general drive" or gen-

[n] The most likely explanation for the data was how the reinforcement variable was originally introduced to respondents. The interviewer began by asking the respondent if he had ever had any of each of 17 very common symptoms. The probability that the respondent had experienced any one of these symptoms at least once in his lifetime was very high. Therefore, a "Yes" answer was most likely a true answer and was reinforced, and a "No" answer was most likely an underreport and was not reinforced. Thus, the concept initially taught was probably, "Give valid answers and don't underreport." This concept may have been maintained by the reinforcement schedule used in the remainder of the interview and was possibly still in effect for questions about chronic conditions.

eral arousal. This refers to a hypothetical concept about a nonspecific or general level of motivation which multiplies the strength of *any* behavior tendency in a person. For example, if a person is running a 100-yard dash, his speed will be faster if his general drive level is high and slower if his general drive level is low. His speed is affected by other things, of course, such as his ability as a runner and his specific desires to win. It is hypothesized that reinforcing statements in the interview increase the respondent's general motivation and thereby accentuate whatever response tendencies exist, regardless of whether the response tendencies are the right ones or not.

A complete presentation of drive theory and supporting empirical evidence is beyond the scope of this section, but one fairly complicated aspect of the empirical evidence is important enough to mention. (For further discussion, the reader should consult Zajonc.[57]) There is a complex relationship between drive and behavior. If a task is easy, high levels of drive will result in good performance. If a task is difficult, high levels of drive interfere with efficient performance while low levels of drive are accompanied by good performance. For tasks of intermediate difficulty, the level of drive shows a curvilinear relationship to good performance, with moderate levels of drive producing good results and high and low levels accompanied by poor performance.

It may be that high levels of respondent drive reduce the underreporting problem in survey interviews and that this has been one effect of reinforcement procedures in the studies already discussed. There is some evidence that leads one to suspect, however, that respondent drive levels may have been too high in some of the experimental interviews. A schematic representation of data from Marquis, Cannell, and Laurent[53] is given below. Possibly, the addition of several



experimental techniques aimed at producing good performance on the moderately difficult recall task (remembering one's chronic conditions accurately) caused performance to suffer in the way the Yerkes-Dodson law predicts.

If an extensive interview, such as the one described by Laurent, had been used in combination with reinforcement instead of long questions, the respondent task would have been easier and accuracy would have been increased.

Effects on specific motives.—Specific motives, such as achievement or social approval, may be affected by reinforcement. The tendency to achieve or seek approval, in addition to being influenced by general drive, can also be heightened or dampened by the amount of social approval given at any particular time. This might be made clearer by the following analogy: If a person is hungry and eats a reasonable quantity of food, he soon behaves as if he is no longer hungry. It is said that the food consumption reduced the desire for additional food. If, on the other hand, a person is not particularly hungry but is allowed to eat a small quantity of food (e.g., one potato chip), he often behaves as if he has become hungrier. Similarly, social reinforcement may increase the tendency to seek social approval (or avoid social disapproval) and task-oriented reinforcement may increase or decrease the desire to do well.

In the Marquis and Cannell study,[50] some evidence was obtained suggesting that the respondent's tendency to seek social approval (or avoid social disapproval) was reduced as a result of receiving reinforcement. These data are far from conclusive, but if the social approval tendency of respondents can be reduced, they have important implications for response accuracy. According to Edwards,[58] people tend to err by responding in a socially desirable direction. If there is some way to make people less concerned about social approval or disapproval in the interview, interview data would presumably be more valid since respondents would be less reluctant to report socially disapproved information about themselves.

It may be that positive reinforcement, which represents social approval given by an interviewer, actually reduces the respondent's tendencies to fear social disapproval and thereby reduces his reluctance to report socially dis-

approved information. Marquis and Cannell[50] also found that reinforced respondents report about one and one half times as many highly embarrassing symptoms for themselves as do nonreinforced respondents. This finding is consistent with the specific motive reduction hypothesis discussed above.

On the other hand, it may be that an interview using only a few reinforcing statements may fall victim to the "one-potato-chip effect." It may be that fear of interviewer disapproval could be accentuated by having only a small number of positive feedback statements programed into the conversation.

Feedback and establishing a relationship.—It is often stated that the success of an interview depends on the degree to which the interviewer can establish rapport with the respondent. Operational definitions of rapport differ and are usually unstated. The concept often seems to refer to a relationship of personal understanding and approval between the interviewer and respondent which is thought to facilitate response accuracy.

The evidence is reasonably clear that feedback or other positive statements which do not necessarily follow any particular contingency schedule result in the respondent approving of the interviewer. What is not clear, however, is whether the resulting positive feeling has anything to do with obtaining valid data.

The relationship between approval (produced by several types of feedback) and performance is found to be nonexistent. Marshall, Marquis, and Oskamp[49] showed that respondents tended to like interviewers who made positive comments about their performance more than they liked interviewers who made only negative comments. However, interviewer comment style was unrelated to accuracy of recall even in these extreme conditions.

Bales' data[59] suggest that, for long-term interactions, feedback statements indicating solidarity, agreement, acceptance, attention, and satisfaction, or promoting the release of tension serve a "maintenance" function. That is, they serve to keep a team or group together and working on a task. Presumably, without this kind of interaction, task-oriented groups would break up without finishing the job at hand. It is not clear, however, that a short-term interview interaction requires this socioemotional kind of interaction to be successful.

Finally, Hyman et al.[60] have questioned the assumption that rapport is desirable in the interview. They point to the possibility that the social relationship implied in a high-rapport interview may prevent the full disclosure of socially unacceptable information. Cannell[61] has proposed an experimental test of this hypothesis.

Thus, while positive feedback in the interview seems to create a good relationship, this relationship per se may have either no effect or negative effects on reporting accuracy. On the other hand, it may be that the relationship somehow interacts with other variables (such as a conditioning procedure) to influence response accuracy. Further research is needed in this area.

EXPERIMENTAL STUDY°

The social reinforcement effect for adult humans is fairly well established. In a review of the literature, Krasner[62] pointed out that reinforcement effects have been demonstrated under numerous settings with a variety of reinforcing statements, with many kinds of responses, and with different types of people. Currently, attention has turned to a consideration of other variables which may be important in producing reinforcement effects on human performance.

One of the major concerns in the preceding discussion was the separation of the instructional or cognitive effects of reinforcement from the more automatic conditioning effects. This issue is very important to survey interview planners for several reasons and has some major implications for theories of human behavior. For example, if the reinforcement techniques mentioned above achieve their effects merely by informing the respondent about what the interviewer wants him to do, it would seem that there are better ways of passing this information along to the respondent. Using only reinforcing statements to convey information about what

°This section was written in collaboration with Ms. Linda Wood, who has undertaken the major responsibilities of design and execution of the experiment described here.

the interviewer wants has the distinct disadvantage that the respondent can misinterpret the message or possibly never get it at all. On the other hand, the reinforcement procedures might be producing better reporting which could not be produced by other means. If this is true, reinforcement procedures should be considered seriously for all personal interviews, as a technique for improving reporting.

This issue of whether the reinforcement effect is purely cognitive or more than that is a major concern of experimental psychologists. These researchers have given a great deal of attention to a process they call "awareness," the respondent's knowledge that reinforcement is being used and that it is used only after he gives certain kinds of answers. A number of researchers[63-65] maintain that the reinforcement effect can be obtained for a human subject only when the subject is "aware" of the response-reinforcement contingency or, in terms used here, is aware of what the interviewer expects him to do. The mere fact that a reinforcing statement follows a particular response is not in itself sufficient to increase the probability of occurrence of that response. An increase can be obtained only when the respondent understands the relationship between his answers and the reinforcing statements. The implication of this position is that one may obtain high levels of respondent performance right at the start of an interview merely by giving clear instructions to the respondent rather than by using reinforcement. Furthermore, since reinforcement takes a longer time to achieve its effects, some respondents may never become aware of what is wanted.

Another group of researchers[35,37,66] maintain that reinforcement can have a direct conditioning effect. They do not deny that awareness or cognitive effects of reinforcement may contribute to the reinforcement effect, but say that these are not necessary in order for the reinforcement effect to occur. The major thrust of the research of the latter group is that reinforcement effects can be produced without the respondent becoming aware of the expectations of the experimenter or interviewer. These studies do not show that reinforcement procedures are more efficient than direct instruction procedures. They do indicate, however, that human behavior can be changed without the necessity of the respondent having to think about the change.

Earlier it has been shown that innovative instruction procedures such as brochures, calendars, and informing the respondent about key questions prior to interviewing were not effective in producing better data. Thus, establishing awareness through these initial procedures was not sufficient to produce desired performance, as Spielberger and others might imply.

However, in view of the large number of studies supporting each point of view, it may be that both interpretations are correct but that each is true under different circumstances. It would seem that there is some other variable that helps to determine whether simple knowledge of what the interviewer expects is sufficient for good responding or whether more complicated conditioning procedures are necessary. A comparison of the two sets of studies suggests that this variable is "task difficulty." Those studies that suggest that awareness is sufficient in itself usually involve a simple task, for example, the selection of a first-person pronoun.[27] Those that suggest that awareness may be unnecessary involve a somewhat more difficult task, such as giving "self-acceptance" responses.[34] The difference in difficulty between these two types of tasks has been obscured by the fact that as long as the interviewer's expectations are unstated, both types of task seem difficult.

It must be hypothesized, therefore, that the conditioning effect of reinforcement brings about changes in respondent ability level or skill in responding adequately (see "Dependent Variables," under "Design of an Experimental Interviewing Approach," this report) and that, to the extent that a respondent's skills are low, the conditioning effects of reinforcement will be greater. That is, the importance of awareness, however obtained, is relevant to knowledge of expectations, but only reinforcement can change skill. When skill needs to be improved, a conditioning procedure is necessary. On the other hand, when the task does not require a high level of skill, direct instructions should produce maximum performance and a reinforcement procedure will not produce any further improvement.

Some of the basic questions that arise in the consideration of reinforcement effects should be

clarified in order to improve the quality of reporting. For example,

a. How important is respondent knowledge of the interviewer's expectations?
b. If this knowledge is essential, is it conveyed better by interviewer instruction, by reinforcement, or by some combination of the two?
c. Is the recall skill of the respondent improved by reinforcement and does the amount of this improvement depend on the difficulty of the task?
d. Does reinforcement provide information to the respondent about the adequacy of his responses and does the importance of this information vary with the respondent's skill?

MEMORY AND INFORMATION RETRIEVAL IN THE INTERVIEW

The purpose of this section is to present some hypotheses about underreporting and to describe an experiment designed to reduce underreporting in a field survey. One of the principal causes of underreporting is the failure of recall; information is not reported because it is not retrieved from memory. Work in interviewing methodology tends to support the assumptions of McGeoch[23] and modern interference theory[24] that information does not disappear from memory but may become difficult to recall because of interfering associations. Only the accessibility of information declines, resulting in a "lessening probability of retrieval from the storehouse."[25] Thus, underreporting is essentially a problem of retrieval, and reporting may be improved by manipulating conditions under which retrieval occurs.

THE INADEQUATE SEARCH HYPOTHESIS

Interviewing methodology indicates that the conditions of recall have a crucial effect upon the outcome of an interview. Some information may be unreported simply because the questions do not convey to the respondent an accurate notion of what information to search for. For example, 19 percent of a sample of respondents interviewed in the Health Interview Survey declared in a followup interview that they thought the interviewer was interested only in "fairly important" things.[55] By itself this observation provides an alternative hypothesis to interpret the underreporting of low-impact items. They may be underreported only because the respondent does not consider them to be relevant and therefore does not even search for them.

Several early experimental studies had suggested that the nonreported material had not been repressed or deeply suppressed, nor had it vanished from memory, but often was simply not elicited by the usual questioning procedures. It seemed likely that the use of different sets of questions and techniques by interviewers could significantly decrease underreporting. For instance, over half the hospitalizations not reported in a first interview were reported in a second interview.[17] An experimental procedure which included a few extra questions, more explanation of purpose to respondents, and a mail followup also resulted in a significant increase in the reporting of hospitalizations.[18] Adding probes to major questions regarding visits to physicians reduced the underreporting by 7 percentage points, from 30 to 23 percent.[1] Another study by Balamuth, et al.[4] indicated that checklists also seemed to reduce the underreporting of chronic conditions. A laboratory study showed that the form of the questions had a great effect on the accuracy of reporting items from a movie.[67] Finally, an experimental study by Marquis, Cannell, and Laurent[53] demonstrated a substantial effect of mere question length upon the validity of the reporting of chronic conditions.

A major contribution of these studies is the demonstration that experimental questioning

strategies can improve reporting by changing the conditions under which the respondent is invited to search for past events. While these studies suggest promising avenues for methodological progress, they are not concerned directly with the cognitive processes involved in recall.

AN INTEGRATIVE HYPOTHESIS: COGNITIVE INADEQUACY OF STIMULI QUESTIONS

Although one can hypothesize that failure to recall information is an important reason for underreporting and also demonstrate that recall can be improved, it is not immediately apparent what steps one should take to improve reporting in the survey interview. Why do customary questioning procedures fail to obtain adequate reporting? Why, for example, if a respondent is asked to report his dental visits of the past 6 months, is he likely not to report them all? A tautologous answer is that for some reason the question was not adequate to stimulate retrieval of information located in memory. The inadequacy of a single question to elicit information suggests a consideration of how information is cognitively organized in memory.

When a person experiences an event, the event is not merely recorded in its original form in the manner of a computer tape, but rather it becomes organized in a perceptual field. As in the old illustration of a blind man describing an elephant, the meaning of an event depends upon how it is perceived and with what other events it becomes associated in memory. Thus, what is a simple, single-dimension variable for the researcher may not be a simple item for the respondent. The dental visit the researcher sees as a simple item of information may be organized in one of several frames of reference by the respondent. The respondent may think of the visit in terms of cost, pain, or loss of time from work. From this point of view the single question about the occurrence of the dental visit may be ineffective; it may fail as a stimulus to recall the event.

As shown in figure 1, the respondent's cognitive organization and the researcher's design of a questionnaire can be viewed as two diverging

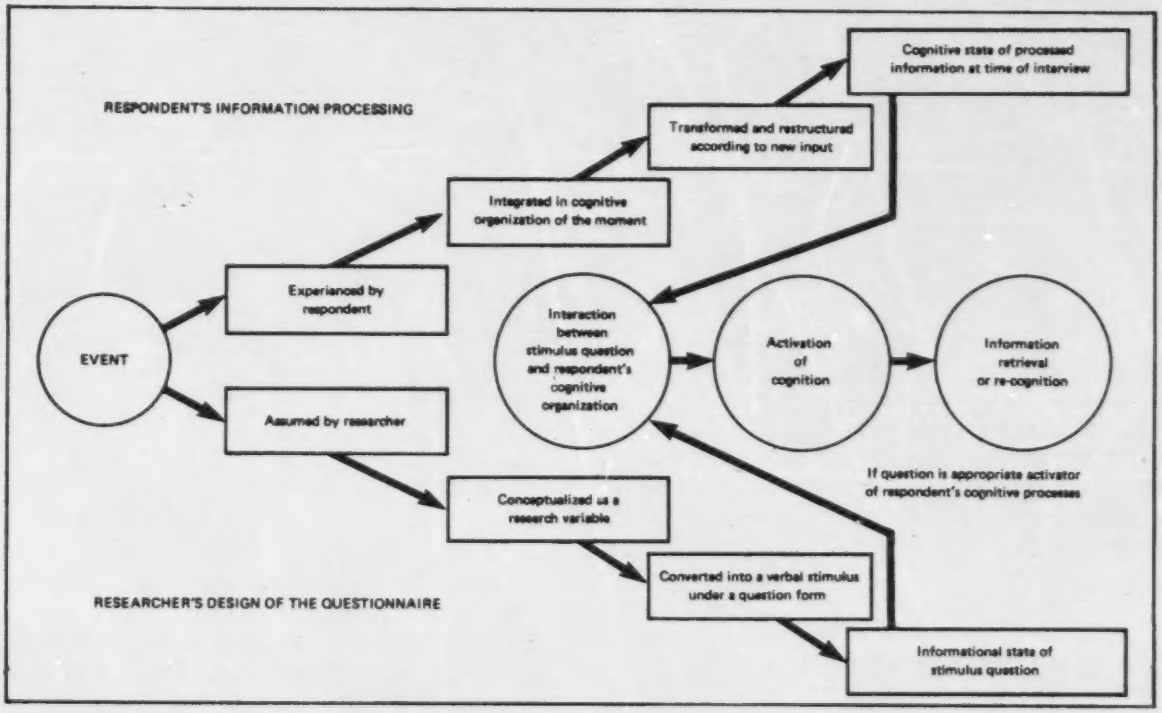

RESPONDENT'S INFORMATION PROCESSING
Cognitive state of processed information at time of interview
Transformed and restructured according to new input
Integrated in cognitive organization of the moment
Experianced by respondent
EVENT
Assumed by researcher
Interaction between stimulus question and respondent's cognitive organization
Activation of cognition
Information retrieval or re-cognition
If question is appropriate activator of respondent's cognitive processes
Conceptualized as a research variable
Converted into a verbal stimulus under a question form
RESEARCHER'S DESIGN OF THE QUESTIONNAIRE
Informational state of stimulus question

Figure 1. Model of information processing in the interview.

paths by which information is processed independently and quite differently before the interview. These two paths lead to two independent informational states—memory trace and stimulus question—whose interaction in the interview is expected to produce the retrieval of the initial information. This model indicates that the probability of proper recall is a function of the ability of the stiumli questions to interact adequately with the respondent's cognitive organization. The appropriateness of the stimuli questions is a function of the researcher's ability to comprehend the nature of the respondent's cognitive path and to utilize this knowledge in the framing of the questions.

The methodological objective becomes one of redesigning questionnaires to facilitate recall. One way of doing this is to incorporate into the design of questionnaires some of the things already known about how people learn, store, and retrieve information. The experiment[p] described below is an exploratory attempt to ascertain the validity as well as the feasibility of this cognitive approach for improving reporting in the household interview.

DESIGN OF AN EXPERIMENTAL INTERVIEWING APPROACH

The study was designed to increase the reporting rate of acute and chronic illnesses, usually underreported in household interviews. The strategy consisted of an experimental questioning procedure which would provide stimuli relevant to the respondent's cognitive processing of health information. This procedure was expected to improve retrieval from memory and to increase the probability of reporting.

To find out about the sicknesses of a person during the month preceding an interview, one may ask a standard question such as, "Were you sick at any time last month?" If the person had had influenza and had stored this experience in memory as "being sick," there is some chance that it would be reported in response to this question.

If, among other illnesses, the researcher is particularly interested in collecting data about influenza, he may also want to use a simple question such as: "Did you have the flu last month?" This question will be a powerful stimulus only if the sickness has been experienced as the flu or influenza and was conceptualized as such. There are clear limitations to a straight application of this recognition technique in the interview. Aside from the old issue of suggestibility, an exhaustive list of all potentially relevant items of information would not usually be feasible. Furthermore, the recognition technique relies much more than other retrieval techniques upon the assumption that researcher and respondent share the same concepts. This is clearly shown in a medical interview. A physician contends that his patients usually report only major surgery when asked about previous operations, whereas they tend to report both major and minor surgery in answer to a question about stitches. The recognition principle is basic to the recall process,[68] but its application to interviewing procedures is valuable only if appropriate stimuli of recognition can be designed.

All of this experience indicates that an event may be stored in memory under various informational states so distant from the initial informational state that a stimulus merely traced from the original event or from its straight conceptualization might not elicit the stored information. Last month's influenza may have been stored in memory in many different ways and not necessarily as being sick or having the flu. There is enough evidence from experiments, as well as from everyday experience, to show that memory is not a simple recording device but rather a complex process in which information is transformed and organized. Thus, influenza may no longer exist in memory as a sickness, but may be organized around a number of other possible traces. For example, the interference of serious chronic illnesses may cause the respondent to fail to consider the episode of influenza as a sickness and no longer to store it as such. On the other hand, this minor flu may have prevented the respondent from going to work on a cold day, thus reducing his income in a manner significant enough to have

[p] A detailed presentation and discussion of this experiment appears in a study by Laurent, Cannell, and Marquis.[69]

impact on his budget. It becomes clear that this respondent may be less likely to report the flu in answer to a single question about sickness than to a question such as: "Have you lost any income because of any sickness during the past month?" If the flu has caused this person to be feverish, a symptom that he very seldom experiences, another relevant stimulus might be: "Have you had a fever during the past month?"

Memory can process an illness in such a way that it gets transformed in storage and becomes organized around concepts such as pain, incapacity, cost, visits to doctors, hospitalizations, medication or treatment, symptoms, or more generally around other causal, circumstantial, or consequential events. The initial perception may be distorted to an association with another perception in order to fit into some structure. If one asks a broad question such as, "Tell me about your illnesses," the respondent has to make the effort to review a multitude of cognitive structures in order to recall properly. He has to invent appropriate frames of reference to guide his search; he has to create his own cues to reactivate traces of possibly weak salience. Altogether, the task is enormous and complex and the motivation to invest substantial effort in it cannot be expected to be spontaneously high, especially in the context of an information-getting household interview that has no immediate benefit for the respondent. Within this framework, underreporting is predictable; the broad question is not an adequate stimulus to the relevant frame of reference and the respondent is not going to work hard to recall information.

This framework, however, provides prospects for an experimental methodology. Instead of asking one standard question essentially traced from a simple conceptualization of the event to be recalled, several questions may be asked that are traced from hypothesized states of information after memory processing. In other words, instead of requesting from the respondent the difficult task of building up his own cues and frames of reference, the researcher should attempt to create these recall aids and to build them into the questionnaire. If the researcher can be successful in predicting and designing the relevant cues and frames of reference, then the respondent's recall process should be significantly facilitated and the availability of information correspondingly increased.

The Extensive Questionnaire

The strategy of an extensive health interview consisted of designing a questionnaire containing a large number of questions that would provide the respondent with multiple and overlapping frames of reference and cues (see Laurent, Cannell, and Marquis[69]). Medical information was asked within classical conceptual frameworks as well as in the language of the layman, and through standard questioning as well as through multiple behavioral cues or direct recognition of items. Transitions between sections were also used to bring some relief in the questioning style and to instill a deliberately relaxed pace in the interviewing.

The questionnaire started with question about symptoms, such as, "Do you have pains in the abdomen?" "Have you had any pain or soreness in your joints?" "Have you had trouble breathing?" Every time the respondent gave a "Yes" answer the interviewer used the probe, "Do you have any idea what causes it?" in an attempt to obtain the report of an underlying health condition. Other frames of reference and cues were used, such as asking for a medical history by means of queries related to childhood, adulthood, 6 or 12 months previous, last week, or the week before last. Then specific behavioral implications of illnesses such as diet, food sensitivity, restrictions of activity, medications taken, and visits to physicians were all used to provide assistance in the retrieval process. Finally, a checklist of chronic conditions implemented a direct items-recognition approach.

Control Questionnaire

A control questionnaire was used in the collection of information on the same major items of health information as the extensive form, but it consisted of single direct questions for each variable. This procedure contained standard questions comparable to those used in the Health Interview Survey and the same chronic conditions checklist that was used in the extensive form.

Field Experiment

The study was designed to compare the effectiveness of the two questionnaires on two main dependent variables: (a) the number of health conditions reported, and (b) the impact of these conditions on the respondent.

To decrease the variance from factors other than those purposely introduced by the experimental design, the sample population was homogeneous—a restricted segment of English-speaking, native-born, white females between 18 and 65 years of age, all of whom were residents of the city of Detroit and were of low-middle and middle socioeconomic status. Two clusters of three dwelling units were chosen at random from each of 110 blocks, selected with probability proportionate to size in 16 census tracts. Only one person in each dwelling was interviewed; the wife in the household was the first choice. Six female interviewers were employed in the study and were assigned to geographically convenient sections of the city. The assignment of experimental questionnaires to households was random within each sample block.

The study showed that 204 respondents were interviewed, consisting of 105 extensive and 99 control subjects. Aside from the variations in questions, the interviewing techniques were kept constant for all subjects (e.g., introduction of the survey, getting demographic information, probing procedures, interviewing style). All respondent demographic characteristics were similar, as were the response rates (87 percent in both interviews). On the average, the duration of the extensive interview was 74 minutes and of the control interview, 40 minutes.

Dependent Variables

Eligible health condition.—To ensure the comparability of the data collected through the two interviewing procedures, precise criteria were established to determine the eligibility of any reported health problem as a condition. Every time any health problem was mentioned in either of the two procedures, standard structured probing was used to ascertain its eligibility as a condition. In order to be an eligible condition for data analysis, an illness had to present either acute characteristics—that is, to have started within the 14 days preceding the interview—or chronic characteristics—to have chronic implications for the person's health. All conditions were screened by the interviewer and later by the coder to include only those meeting the criteria of the Health Interview Survey. The number of eligible conditions reported was the major dependent variable analyzed.

Condition impact.—As suggested earlier, health events get transformed and organized in memory within various clusters or frames of reference. It is assumed that the number of such categories under which a health condition is organized is a mark of its salience or impact and a predictor of its accessibility for retrieval. A condition of low impact that is organized under only a few categories is likely to be missed by a single question. Therefore, the hypothesis in this study was that the extensive interview would pick up more conditions of low impact.

To create a measure of impact in both techniques, every eligible condition reported was probed according to a standard procedure. Additional information was asked about the existence of health behaviors associated with each condition occurring during the past 2 weeks (visits to a physician, medicine, treatment, special diet, days in bed, days of restricted activity, amount of pain or discomfort). An index of impact was created for each reported condition on the basis of this additional information.

The two dependent variables were selected upon the empirical evidence that underreporting represents a major problem in the health interview. As medical records were not available for this research, a working assumption built into the design of the study was that the more information reported, the better the overall validity.

Hypothesis

The major hypothesis that the study attempted to test was the following: By providing a broad aid to the respondent's recall process through the use of multiple frames of reference and cues, the extensive procedure is expected to increase the overall number of reported eligible conditions as compared with the number obtained in the control procedure. Since events of low impact are more likely to be underreported than those of high impact, a significant increase

was particularly expected in the reporting of chronic conditions of low impact within the extensive interview. As a consequence, the overall impact level of the reported conditions was expected to be lower in the extensive than in the control procedure.

Results and Discussion

As shown in table 38, the data clearly supported the major hypothesis. A significantly larger number of health conditions was elicited by the extensive interview than by the control interview. Respondents reported an average of 7.88 eligible conditions in the former and 4.42 in the latter. The increase in reporting was significant for both chronic and acute conditions.

As predicted, the increase in reporting was obtained by eliciting conditions of low, but not trivial, impact that were not reported in the control interview. As shown in table 39, the mean level of impact for all conditions was significantly lower in the extensive procedure (2.03) than in the control procedure (2.64), and this was especially true for chronic conditions. It was further verified that the reporting of higher impact conditions was similar for both techniques.

Several analyses have been conducted in an attempt to identify the main sources of improved reporting in the extensive questionnaire. It was found, for instance, that 61 percent of all conditions reported in the extensive interview were elicited by the initial symptoms questions. These questions dealt with the presence of specific symptoms, aches, or pains and with the health conditions underlying them. The average number of conditions reported under these initial questions in the extensive interview was larger than the total average number of conditions reported in the entire control interview. This observation indicates that a cue-giving approach using symptomatic manifestations of illnesses as a frame of reference is more produc-

Table 38. Mean number of health conditions reported per person and difference between means, by type of condition and questionnaire procedure

Reporting variable	Mean number of conditions reported		Difference between means	p^1
	Extensive procedure (105 respondents)	Control procedure (99 respondents)		
All eligible health conditions	7.88	4.42	3.46	.001
Chronic conditions	6.29	3.99	2.30	.001
Acute conditions (in last 14 days)	0.82	0.33	0.49	.001

[1]Significance level of difference. Values were computed on the basis of the one-tailed *t*-statistics.

Table 39. Number and mean impact level of health conditions reported, by type of condition and questionnaire procedure

Reporting variable	Number of health conditions		Mean impact level		Difference between means	p^1
	Extensive procedure	Control procedure	Extensive procedure	Control procedure		
All eligible health conditions	661	399	2.03	2.64	−0.61	.001
Chronic conditions	583	370	1.87	2.46	−0.59	.001
Acute conditions (in last 14 days)	73	29	3.34	4.93	−1.59	.025

[1]Significance level of difference. Values were computed on the basis of the one-tailed *t*-statistics.

tive in eliciting the report of illnesses than are the standard questions actually designed to serve this purpose.

Another observation of this kind concerns the effectiveness of recognition checklists. At the end of both interview procedures, a standard checklist of 41 chronic conditions was utilized and the relative effectiveness of this device was evaluated within each procedure. It was found that this last section was highly productive in the control questionnaire (57 percent of all conditions were first reported there) and appreciably less productive in the extensive questionnaire (where it yielded a total of 16 percent of first reports). The high figure obtained under the control procedure (57 percent) confirms to some extent the effectiveness of an items recognition approach as a basic tool for eliciting information in a standard type of interview. The lower but still substantial figure obtained under the extensive procedure (16 percent) demonstrates the adequacy of this recognition approach for gathering information not already reported through the various cues provided during the course of an extensive interview.

A comparison of the differential effect of this recognition approach under the extensive and control interviews for the reporting of chronic conditions is shown in table 40. In the extensive interview, 67 percent of the chronic conditions in the final standard checklist were first elicited before the final checklist, as compared to only 26 percent in the control interview. In both techniques, the high-impact conditions were reported first, the low-impact ones last. This is especially typical in the control technique where the mean impact level varies from 4.36 for chronic conditions reported prior to the recognition list to 1.61 for those reported on the recognition list. An analysis of variance carried out on the impact level of all eligible conditions reported by chronological sections of the questionnaires in the two procedures led to the same observation, showing a significant lowering of impact from the earlier sections of the questionnaires to the later sections.

The same finding was replicated on a question basis in the extensive interview. The extensive interview made some use of primary or standard questions followed by additional or cue-giving questions. The average impact level of eligible conditions reported for each of these two types of questions is shown in table 41. An analysis of variance based on these data showed that the average impact score varies significantly ($p \leq .05$) according to whether the condition was reported in the primary or additional questions. Table 41 demonstrates the tendency for the average impact to be lower for conditions reported on the additional questions (2.25) than on the primary questions (3.66). This table also illustrates clearly the power of additional cue-giving questions to elicit information not reported on standard questions. In this case the additional questions are providing as many and even more eligible conditions (36) than the primary ones (32).

This general tendency for high-impact conditions to be reported earlier and for low-impact conditions to be reported later—within an entire questionnaire or within given questions—confirms the idea that high-impact events are more easily recalled and reported than are low-impact events. Since they are easier to

Table 40. Percent of chronic conditions reported and mean impact level of listed chronic conditions, by whether first reported prior to or in response to recognition list, by questionnaire procedure

Questionnaire procedure	Listed chronic conditions				Total listed chronic conditions reported	
	First reported prior to recognition list		First reported in response to recognition list			
	Percent	Mean level of impact	Percent	Mean level of impact	Percent	Mean level of impact
Extensive	67	2.29	33	1.48	100	2.02
Control	26	4.36	74	1.61	100	2.32

Table 41. Number and impact level of eligible conditions, by whether first reported in primary or additional questions of the extensive questionnaire

Questions in which condition was first reported	Number of conditions	Impact level	
		Mean	Standard deviation
Both types of questions	68	2.91	2.36
Primary questions	32	3.66	2.65
Additional questions	36	2.25	1.87

NOTE: $F = 6.50$ ($p < .05$).

recall, they are reported first; low-impact items appear harder to obtain from the respondent and, as such, require stronger stimuli (recognition lists or cues) and more time. Thus the recent behavioral implications, or impact, of an event strongly affect the likelihood of its retrieval. This likelihood seems to increase as the amount of behavioral implications increases, and vice versa.

The inadequacy of standard interviewing to elicit reports of lower impact events appears, then, as a major cause of underreporting and incomplete data. The analysis presented above has pointed to the ability of some interviewing devices to improve the prospects of reporting for events whose behavioral implications are weak. It is interesting to note that in spite of its low-impact character, the additional information collected through the extensive technique was still significant in terms of the criteria of the Health Interview Survey, as it involved, for example, restricted activity or use of medical services.

Conclusions

The results of this study emphasize the important effect of question content and question strategy on survey data and suggest several methods of reducing underreporting bias. The methodological objective was to demonstrate that the completeness of the reported information can be changed significantly by programing the respondent's recall task more efficiently. Standard questions may not represent the most adequate stimuli to activate respondent recall because they may ignore the way in which information is organized in memory. Thus an attempt was made to pattern an experimental questioning procedure after the processes that the respondent was expected to use in acquiring, storing, and retrieving information. This was accomplished by the use of multiple frames of reference and multiple cues integrated into a questionnaire. A substantial increase in information was obtained through these procedures in the areas of information where underreporting is traditionally observed. This improvement is interpreted as the result of a greater correspondence between the questioning procedures and the manner in which respondents organize health information in memory.

Several uncertainties remain, however. While there is evidence from the reported data that the additional information elicited in the extensive procedure is not trivial, the validity of this information needs to be ascertained by a study which would check the respondent's reporting against valid records. The amount of overreporting obtained with the experimental technique should be evaluated and compared with the amount obtained using a control technique.

On a more theoretical level, although the results of this study supported the hypotheses, it is not possible to infer from these data any satisfactory statement of causality. Indeed, the interactions between cognitive and motivational factors involved in the interview situation were not controlled in this study. Even though the approach was cognitive, it seems that motivational changes also occurred that may have been instrumental in determining the outcome of the experimental treatment. For instance, as the extensive procedure made the recall task easier, it also reduced the amount of effort required to perform it and thus reduced the motivational requirements of the task as well. Also, since the interviewer had to devote more time and display more behavior in administering the extensive questionnaire than in administering the control procedure, this increased activity may have conveyed the idea that the task was important, and may have thus heightened the respondent's motivation to perform. Too, the respondent may have modeled his behavior after that of the interviewer (the interviewer did more talking in

the extensive interview and the respondent may have followed his lead by being more active and reporting more information). It is unclear whether increases in reporting have been obtained through direct cognitive facilitation, reduction of motivational requirements, indirect motivational stimulation, or a combination of these factors. The major outcome was a pragmatic one; techniques designed in a cognitive framework to facilitate recall have proved effective in increasing reported information.

Finally, two main implications deserve particular attention. First, it seems that the complexity involved in retrieving adequate information in an interview tends to be underestimated. Even data that appear relatively simple to obtain may require the design of more elaborate interviewing techniques. Efforts have been made in past research to overcome various biases in interview responses. However, in most cases they have been avoidance strategies rather than constructive approaches. Evidence from data such as those presented above suggests that probably very little is known about the asking of appropriate questions, so that reporting errors may often be the result of questioning errors. In view of the wide utilization of survey methods to create new knowledge and to guide policies, there is great need for more basic research on the interviewing tool itself. A major objective for these research efforts is simply improvement in the design of questions. The present research has attempted to demonstrate that progress can be made in this direction by framing questions in accordance with the respondent's cognitive processing of the initial information.

A second implication of this research is that the experimental survey research interview could provide a new approach for investigations in the field of cognitive psychology. If hypotheses about the cognitive process can be introduced into the design of interviewing experiments, then the interview setting can serve as a laboratory for the study of human memory and recall. It is interesting to note that most of the knowledge related to the memory process has been developed in classical experiments by manipulating the input or learning conditions and evaluating the resulting output or recall. From this experimental design, inferences are made about cognitive processing and memory. Thus, in the laboratory the focus is most often on input or learning conditions. Little attention is given to recall, which is usually considered as an end result variable or as a test of learning and retention after memory processing. Textbooks are more likely to discuss the psychology of learning than the psychology of recall. A reversed strategy was attempted in the present research; the learning conditions or input were kept constant or controlled by experimental survey design and conditions of recall were manipulated. Recall was no longer considered only a test of learning but was viewed as a powerful intervening process itself—one which mediates the effects of learning and memory processing on survey interview reporting.

"The more questions one asks about a topic, the more information one obtains," is a statement made frequently by survey researchers. "Recognition-list questions will obtain more information than those requiring free recall," is commonly stated by survey practitioners. The present study helps to provide some understanding of the phenomena underlying these statements. It is not a simple matter of asking more questions; nor are recognition-list questions necessarily better than other kinds of questions. Some basic principles of memory and retrieval can be used to improve reporting. This study suggests that further research will yield a better understanding of the way in which information is stored and will invent more effective methods of retrieving that information.

QUESTION LENGTH AND REPORTING BEHAVIOR IN THE INTERVIEW: PRELIMINARY INVESTIGATIONS

This section reports on two exploratory studies of the effects of question length on the amount and validity of information reported in household interviews. Both of these studies are described in more detail in *Vital and Health Statistics*, Series 2-Number 45,[53] and the findings of the second one have been discussed previously in this report in relation to the effects

of reinforcement on household interviews in the section, "The Use of Verbal Reinforcement in Interviews and Its Data Accuracy."

Statements such as the following can be found in questionnaire methodology manuals: "Make the questions as concise as possible .... The length alone makes it practically impossible to carry the question as a whole in mind."[70] "It is generally best to keep questions short—preferably not more than 20 words."[71] On the basis of these statements it was assumed that lengthy questions were an obstacle to clear communication. However, some empirical findings showing the presence of a verbal behavioral balance in the household interview[65] along with other findings repeatedly demonstrated a matching effect between interviewer and respondent speech duration.[72] From this finding it was concluded that long questions might have some value not previously anticipated.

Empirical Findings on Behavior Matching In the Interview

The Survey Research Center study by Cannell, Fowler, and Marquis[q] showed a clear, positive association between the overall amount of behavioral activity of the respondent and the number of items reported. Furthermore, a very high correlation was found between the behavior activity level of the interviewer and that of the respondent. These findings led to speculation that interviewer and respondent sought and perceived cues from each other about the degree of effort to put into their respective roles. If this cue-search process causes the respondent to model his behavior after the behavior of the interviewer, inferences can be made about question length. Logically, short questions should elicit short answers and long questions should yield long responses.

This inference is supported by a series of studies on interview speech behavior conducted by Matarazzo et al.[73] Briefly stated, these studies demonstrated that some formal measures of the interview process unrelated to content—namely, interviewer and respondent speech duration and silences—are remarkably reliable, valid, and consistent. These studies have shown very explicitly and repeatedly in employment interviews that an increase in interviewer average speech duration resulted in a significant increase in respondent average speech duration. For instance, in a 45-minute interview divided into three 15-minute periods where the interviewers spoke in utterances averaging 5.0, 15.2, and 5.5 seconds, the respondents' utterances averaged 30.9, 64.5, and 31.9 seconds, respectively. In other experiments of this series, the researchers varied the schedule of the interviewer sequence of utterances both in range and direction. They also controlled for number and type of questions, for topics discussed, and for interviewer differences. In all cases they consistently obtained changes of approximately 100 percent in respondent speech duration. The changes were always in the direction of the patterns shown by the interviewers.[r]

The results of these studies of behavior matching in the interview demonstrate that increases in interviewer verbal activity produce increases in respondent verbalization. The probability that a long response will contain more information is an interesting hypothesis to be tested. The research described here is devoted to an investigation of this problem.

Hypotheses About the Effects of Question Length on Reporting Behavior

A major contribution of the work by Matarazzo and his associates was the focus on the measures of speech behavior in the interview not related to content, particularly interviewer and respondent speech duration. Survey researchers are interested in the potential effect of these variables (question length or speech duration) upon the content variables (amount and validity of reported information). Ways of inducing greater respondent verbalization are of particular interest because it might be conducive to improving reporting accuracy.

---

[q] This study is described in greater detail in an earlier section of this report, "Behavior in Interviews." For a full report of the study, see reference 46.

[r] Replications of this finding have been obtained under other interviewing situations, such as the astronaut-ground communicator conversations,[73] the Kennedy news conferences,[74] and the experiment on interviewer style by Heller, Davis, and Myers.[75]

One needs to ascertain if the speech-matching effect is present in survey interviews. Furthermore, it is necessary to find out whether the increases in respondent speech duration reported in other studies are a direct result of increases in interviewer speech duration or the result of greater information demands made upon the respondent by more elaborate questions. If person A is asked the short question "Tell me a little bit about your job," and person B is asked the long question "Tell me everything you can think of about your job. I am interested in as many details as you can provide to describe it," one expects person B to talk longer than person A simply because of differences in the demand characteristics of the question. In order to find out whether changes in respondent speech duration are a function of changes in interviewer speech duration and independent of changes in the information demands of the question, one needs to modify the length of the question while holding constant the amount of information it requests. Such a design was used in the first experiment to be described in this section.

A second crucial task is to find out whether increases in question length, without changes in the information demanded, have any effect on the answer content, independent of variations in respondent speech duration. A mere increase in question length might result in an increase in amount of reported information, with or without a corresponding increase in respondent speech duration. This hypothesis is based upon the cue-search model of the interview described above, in which it is assumed that the respondent looks to the interviewer as a source of cues. A long question might provide the respondent with cognitive and motivational cues conveying the idea that a full report is desired. It also gives the respondent more time to work on recall while the long question is being asked. The final reporting performance might change, although this change might not necessarily be reflected in the length of the answer.

Finally, it is important to ascertain the effect of question length upon the validity of the reported information. Validity might be improved by the use of long questions, since the interviewer cues might transmit a request for completeness and accuracy of report. For this step of the research, the respondent's report will be checked against independent records.

EXPERIMENT 1: EFFECTS OF QUESTION LENGTH ON ANSWER DURATION AND REPORTING FREQUENCY

A pilot field experiment was designed to test the effects of interviewer speech duration upon respondent speech duration and reporting frequency.[53] In order to increase control over all aspects of interviewer verbal behavior, variation in speech duration was created through the use of questionnaires with short and long questions. Furthermore, the lengthening of short-form questions was implemented in such a way that the information requested in short and long questions was kept constant. This strategy was used to rule out the possibility of obtaining longer answers to longer questions only because these questions explicitly asked for more information.

Questionnaire Procedures

An interview containing 28 questions was created according to customary methodological principles. An average of 14 words was used in each question. Information was requested on various health events (e.g., acute illnesses, injuries or accidents, and chronic conditions) and health-related behaviors (e.g., medicines taken and doctor visits) which occurred during various periods of time (e.g., last 2 weeks, 4 weeks, 6 months). The questions were primarily about the respondent herself (24 questions) and secondarily about another selected member of the household (4 questions). The types and numbers of questions used were as follows:

a. Open-ended type, leading to free response (8);
b. "How many . . ." type, seeking information on frequency of specific events (8); and
c. Closed, forced-choice, checklist type, dealing with presence or absence of specific chronic conditions (12).

Then, each question was written in a long form according to the following procedure. A long question consisted of three sentences:

a. An introductory statement giving a partial description of the topic of the question, including the same terms as used in the short question, but in a different grammatical structure, possibly using a cliché;

b. An intermediary statement conveying information already contained in the short question but not presented in the introductory statement, and usually introduced by another cliché; or a filler, introducing some extraneous information of obvious and inconsequential nature about the survey but unlikely to affect the meaning of the question; and,
c. The question itself in its short form.

The two following examples provide an illustration of this question-writing procedure:

Q. 4. *Short Form:* Would you tell me what accidents or injuries you may have had during the last six months? (17 words)

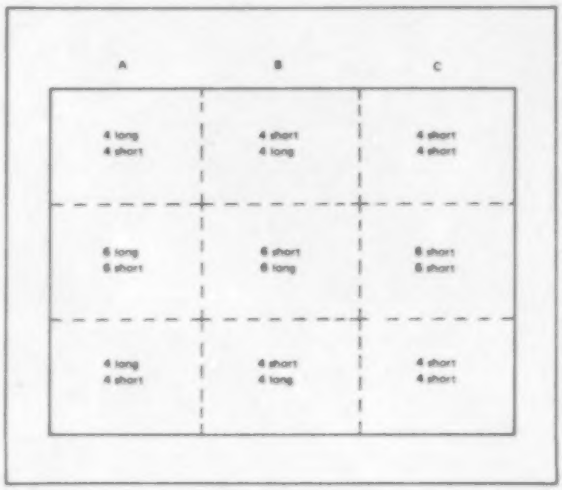
Q. 4. *Long Form:* We would like to go next to a question about accidents and injuries. In this survey we ask everybody to report their accidents and injuries for the last six months. Would you tell me what accidents or injuries you may have had during the last six months? (47 words)

Q. 17. *Short Form:* Have you ever had any trouble hearing? (7 words)

Q. 17. *Long Form:* Trouble hearing is the last item of this list. We are looking for some information about it. Have you ever had any trouble hearing? (24 words)

Thus, length was added to questions by introducing redundancy, clichés, and extraneous information. It was assumed that this procedure did not alter the objective or meaning of the question. The short form of the question always appeared with identical wording in the last part of the long question. Long questions contained an average of 38 words each. They were 2.7 times longer than the average short question. This length ratio was assumed to be large enough to ensure a substantial variation in interviewer speech duration, despite expected variations in the speed of reading.[5]

Three questionnaire procedures (A, B, and C) were designed from a pool of 28 questions, all written in both short and long forms. The 28 questions were used in the same order in all 3 procedures. Questionnaire C (control) consisted of short-form questions only. Questionnaires A and B consisted of blocks of long-form and blocks of short-form questions, alternated in such a way that each block of questions asked in the long form in procedure A was asked in the short form in procedure B and vice versa. This particular arrangement was used for two reasons. First, it was assumed that a questionnaire employing only lengthy questions might be detrimental to useful respondent performance.[f] Second, this particular design allowed for an investigation of potential carryover effect of length from blocks of long questions to blocks of short questions.[u] Finally, this design allowed for a comparison of answers to all questions in short form in treatment C and long form in either treatment A or B (with a mixture totaling about one-third short questions and two-thirds long questions). Thus, the three experimental procedures presented the following question-length composition:



Field Procedures

Two female interviewers on the Survey Research Center staff each received about 8 hours

[5] Ray and Webb's research on Kennedy news conferences[74] used the number of lines of transcripted speech as the unit of speech duration analysis. That measure has been found by Matarazzo[72] and others to be highly correlated with standard methods of time recording (over .90) and highly reliable.

[f] Current Survey Research Center work tends to discard th assumption, at least for interviews of moderate total length (about 15 minutes).

[u] The carryover effect was not found by Matarazzo, but was suggested by the Ray and Webb study of Kennedy news conferences.

of training. They were instructed to read the questions exactly as worded on the questionnaires, to avoid any obtrusive speech behavior, and to provide clarification by repeating only the relevant part of the question and only when absolutely necessary. They were also told to omit probing, to accept the respondent's answer, and to eliminate any kind of verbal or nonverbal feedback. Furthermore, they were trained to adopt a regular speech rhythm, consistent for questions of varying length. They were told that the purpose of the study was to experiment with various types of question structure.

Four city blocks were selected at random from two census tracts in Jackson, Michigan. The tracts contained white families of moderate income, with a high proportion of native-born citizens and a low proportion of persons over 65 years of age. Two blocks were assigned to each interviewer and interviews were taken according to a random selection of dwelling unit numbers. In order to be eligible, respondents had to be white, 18 to 64 years of age, married, female, able to respond adequately, and fluent in English.

A total of 27 interviews were taken (9A, 9B, and 9C). Questionnaire forms were randomized among the four blocks, both in terms of number and administration order. Interviewer 1 took 4A, 5B, and 5C forms; interviewer 2 took 5A, 4B, and 4C forms. All interviews were tape recorded with cassette-type machines.

Dependent Variables

Two dependent variables, assumed to be affected by the length of the question asked, were measured:

a. The duration of respondent answers to each question. This measure was defined as the number of seconds from the end of the question to the end of the response minus any irrelevant interruption or any additional interviewer verbal intervention.[9]
b. The percentage of questions in which one or more items of the requested health information were reported.

[9]Answer duration was timed from the tapes by a single coder using electronic timers.

Results and Discussion

Answer duration.—The average number of seconds the respondent took to answer a question in relation to question length is shown in table 42. Two converging findings emerge. First, it is clear that within interviews containing both short and long questions (A and B) the long questions did not elicit any longer answers than the short questions did (5.6 seconds against 5.7 seconds). Second, interviews with only short questions (C) did not elicit substantially shorter answers than long questions did in the other interviews (A and B). The average answer length is slightly lower for short questions used alone (5.3 seconds) than for long ones used in combination (5.6 seconds), but the difference is not statistically significant and may be considered inconsequential.

Table 42. Number of respondents, number of questions asked, and average duration of answers per question, by question length

Question length	Number of respondents	Total number of questions asked	Average duration of responses per question (in seconds)
Long questions in interviews using both long and short questions (questionnaires A and B)	18	270	[1]5.6
Short questions in interviews using both long and short questions (questionnaires A and B)		270	[1]5.7
Short questions in interviews using short questions only (questionnaire C)	9	270	[1]5.3

[1]Differences between these figures are not statistically significant (two-tailed *t*).

These results are far from the approximately 100-percent increase in answer duration repeatedly obtained in other research where comparable lengthening of interviewer speech had been used. In the limited exploration under-

taken, the matching effect between interviewer and respondent speech duration did not appear. It could be argued that in contrast to other types of interviews which usually include many open questions, a survey interview containing a substantial number of closed questions does not provide a chance for respondents to do much talking; therefore, the matching effect does not emerge.

To provide some understanding of this issue, a further analysis of answer duration was conducted which controlled for question type (open or closed). While the average answer duration was longer for open questions than for closed questions, the matching effect of question length was not more apparent for the former than for the latter. These results do not replicate the Matarazzo findings described earlier.

In contrast with other research, the major change introduced in this question-type experiment was a strict control on the information load and demands of the questions, so that a long question would not transmit or ask for more information than would a short question. In view of the results and pending replication, it is very possible that this control was responsible for the discrepancy between the present data and the Matarazzo findings. The absence of a length-matching effect under the new experimental conditions supports the authors' interpretation of the findings by Matarazzo and others, proposed earlier as hypothesis. Variations in respondent speech duration may be quite independent of variations in interviewer speech duration, resulting instead from changes in information content and demand of a question. When information level is held constant across short and long questions, there is no longer any clear indication of a matching effect of speech duration.

Reporting frequency.—The effect of question length upon the probability that responses contain one or more items of requested health information is shown in table 43. It is apparent from the table that within interviews containing both short and long questions, the length of a question does not seem to predict the frequency of report. Thirty-eight per cent of the short-form questions elicited information, compared to 40 percent for the same questions written in a long form. The difference is inconsequential. On the other hand, in interviews using only short questions, only 29 percent of the questions elicited some health report. This proportion differs significantly ($p \leq .05$) from the proportions obtained when both short and long questions were used in the same interview.

Table 43. Number of respondents, number of questions asked, and percent of questions for which any requested health information was reported, by question length

Question length	Number of respondents	Total number of questions asked	Percent of questions with information reported
Long questions in interviews using both long and short questions (questionnaires A and B)	18	252	[1]40
Short questions in interviews using both long and short questions (questionnaires A and B)		252	[1]38
Short questions in interviews using short questions only (questionnaire C)	9	252	29

[1]Both of these proportions are significantly different from the proportion obtained in the third group ($p \leq .05$, one-tailed, based on Z).

Thus, the lengthening of half the questions in questionnaires A and B is responsible for a significant increase in the frequency of report, compared with the frequency obtained in the control questionnaire C. Further analysis of the data by question type showed this effect to be present in answers to both closed and open questions.

Two suggestions emerge from the data shown in table 43. First, the probability of report may be enhanced by the use of longer questions. Second, this effect may be carried over from long to short questions when blocks of both types are used in the same interview. Thus, while there seems to be a positive effect of question length per se upon the probability of report, this effect also appears throughout the entire interview.

In summary, the results of this pilot experiment offer the following proposition. When information demand is held constant, long questions do not produce noticeably long responses, but do elicit a greater frequency of report. Under these circumstances, increases in interviewer speech duration affect the content of the respondent speech without affecting its duration. Somehow a long question provides cues which tend to elicit more information from the respondent even though the response duration stays practically unchanged. Tentative explanations of this question length effect will be proposed in the conclusions to this section, along with an experimental design for the test of some derived hypotheses.

EXPERIMENT 2: EFFECTS OF QUESTION LENGTH ON VALIDITY OF REPORT

While there is some evidence from experiment 1 that increases in question length result in greater reporting frequency, no data are available so far on the validity of the extra information obtained by long questions. It may be that respondents overreported a number of health events or conditions, as is known to occur sometimes in health interviews. However, there is no particular hypothesis to predict that overreporting would increase as a function of question length. On the contrary, drawing again from the cue-search model of the interview, it is hypothesized that respondents may interpret long questions as calling for completeness and accuracy of report. This would decrease underreporting, and possibly overreporting, while improving the overall validity of the data.

The purpose of experiment 2[w] was to ascertain the validity of health information reported in response to long questions. To achieve this objective, a sample of respondents was drawn from a population of patients who had visited a physician in a prepaid clinic during a 6-month period prior to the survey. At the time of the visit, the physician was asked to survey a checklist of 13 chronic conditions indicating whether the patient had or did not have each listed condition, or whether sufficient diagnostic information was available. Information about the patient was obtained by the physician from the patient's record and from his own knowledge of the patient's health. A weighted sample of persons was used in which 88 percent had at least one of the listed chronic conditions and 12 percent had none of them. Respondents were white females, 18 to 60 years of age who resided in the greater Detroit metropolitan area.

Among other experimental techniques used in the study was a test of the effects of question length on accuracy of reporting. A questionnaire was prepared using standard short questions. These call for the reporting of various health conditions and behaviors. In the middle of the questionnaire, checklist-type questions were introduced which asked about the presence or absence of the 13 chronic conditions listed on the physician summary form.

A second version of the questionnaire was prepared using the same questions but in long form. Questions were lengthened in a way somewhat comparable to the method used in experiment 1. Since experiment 1 had shown a carryover effect of question length from long to short questions, it was hypothesized that contrast between questions of various lengths rather than specific question length per se was the operating variable. In order to increase this contrast effect, a mixture of questions varying in length, rather than the previous large blocks of short and long questions, was used. Also, in contrast to the procedure used in the first experiment, standard short probe questions were introduced after all items in both questionnaire forms. This meant that the more information a respondent reported, the more short probe questions she would be asked. This procedure was aimed at decreasing the contrast in length between the two questionnaire forms as a function of reporting frequency, thus minimizing the expected effect of question length on the number of items reported. Finally, the strategy used in lengthening the questions was slightly different. Some questions contained three statements as in the original experiment, while others contained only two.

Thus, while question length was implemented on the basis of comparable principles (redundancy, clichés, and fillers with the information characteristics of the question held constant),

[w] For a full report of this experiment, which also included an investigation of the effects of verbal reinforcement and reinterview, see reference 53.

less differentiation was probably attained between the two experimental treatments. This might result in some minimization of the question length effect.

Ten white female interviewers were employed in the study. Question treatments, long and short, were assigned at random within geographic clusters of respondents with each interviewer administering both types of interviews. Budget considerations led to some compromise, both in terms of experimental design and sample size, which might be reflected in less than optimal stability of the results. One hundred and six persons were interviewed with the short-question procedure and 96 with the long-question procedure.

Since, for various reasons, errors can exist in the physician forms as well as in the respondent reports, a high degree of agreement between the two sources was not expected. However, it was assumed that better validity of the respondent reporting would increase the agreement rates between the two sources. In other words, if agreement rates on presence and absence of chronic conditions were found to be higher under long-question interviews than under short-question interviews, presumably an improvement in validity had been obtained in the former procedure.

Probability of agreement between physician and respondent on the presence or absence of the listed 13 chronic conditions was computed on the basis of match and mismatch in "Yes" and "No" responses provided by the two information sources. Excluding cases where the physician or respondent lacked sufficient information to determine the presence of a condition, the following four possibilities of match or mismatch existed for each chronic condition:

		PHYSICIAN	
		Yes	No
RESPONDENT	Yes	A	C
	No	B	D

Results and Discussion

The original report of the study treated the data in terms of two types of mismatch:

$$\text{Type X} = \frac{B}{A+B} \text{ and type Y} = \frac{C}{C+D}$$

Type X mismatch was expected to represent the extent of underreporting (false negative responses); type Y, the extent of overreporting (false positive responses), assuming that the physician was right in his evaluation. According to the analysis, the overreporting error was relatively infrequent and was not affected by the question length (.10 under both interview techniques). On the other hand, the underreporting error was more frequent and it was reduced significantly ($p \leq .05$) in the interviews in which long questions were used (.46 with short questions and .38 with long questions).

Since one may reasonably question the adequacy of the physician's report, as well as the validity of respondent report, an analysis of the data from a slightly different point of view is presented. For both interview treatments, table 44 shows probabilities of agreement relating to the *presence* of chronic conditions.

Table 45 presents the corresponding probabilities of agreement on the *absence* of chronic conditions.

This new approach avoids the assumption that the medical records were more valid than the respondent's report. It does assume that greater agreement between the two sources indicates higher validity of the reported data.

Based on the material shown in table 44, whether one starts from the physician data (row 1), or from the respondent data (row 2), or from both (row 3), the probability of agreement on the existence of a chronic condition is consistently higher with interviews using long questions than it is with interviews using short questions. For two agreement rates the obtained improvement is statistically significant. The increase in overall probability of agreement obtained with long-question interviews amounts to 16 percent.

The data in table 45 show that while agreement rates on the absence of chronic conditions are not noticeably enhanced by the use of long questions, neither are they diminished by question length.

Table 44. Probability of physician-respondent agreement on the presence of chronic conditions, by type of agreement rate and question length (original interview)

Type of agreement rates	Questionnaire procedure[1]		Difference	Percent increase due to question length
	Short questions	Long questions		
$\frac{A}{A+B}$ Probability that chronic conditions checked as present by the physician were reported as present by respondent	.537	.622	[2].085	+17
$\frac{A}{A+C}$ Probability that chronic conditions reported as present by the respondent were checked as present by the physician	.477	.516	.039	+8
$\frac{A}{A+B+C}$ Overall probability of agreement between physician and respondent for chronic conditions mentioned as present by either of them	.338	.392	[3].054	+16

[1]Number of persons interviewed were 106 in short-question procedure and 96 in long-question procedure.
[2]$p < .05$, one-tailed, based on Z.
[3]$p < .10$, one-tailed, based on Z.

Table 45. Probability of physician-respondent agreement on the absence of chronic conditions, by type of agreement rate and question length (original interview)

Type of agreement rates	Questionnaire procedure[1]		Difference	Percent increase due to question length
	Short questions	Long questions		
$\frac{D}{D+C}$ Probability that chronic conditions checked as absent by the physician were reported as absent by the respondent	.901	.901	0	0
$\frac{D}{D+B}$ Probability that chronic conditions reported as absent by the respondent were checked as absent by the physician	.921	.934	.013	+1
$\frac{D}{D+B+C}$ Overall probability of agreement between physician and respondent for chronic conditions mentioned as absent by either of them	.836	.847	.011	+1

[1]Number of persons interviewed were 106 in short-question procedure and 96 in long-question procedure.

This information provides an interesting complement to the earlier findings. Experiment 1 has shown that more information is elicited by long questions than by short questions, and the present experiment indicates that long questions also elicit information of higher validity.

Even though the study was designed primarily to test the validity of reporting of chronic conditions rather than simply the amount reported, attention was also given to the number of reported conditions. From the list of 13 chronic conditions, an average of 2.26 were accounted for in the long-question treatment as compared to 2.06 in the short-question treatment. Although this amounts to only a 10-percent increase in the number of chronic conditions reported, it is not negligible in view of the fact that the recall task is easier and the likelihood of responding is expected to be uniformly higher when recognition stimuli like checklist questions are provided.[53] For health behavior noted in other parts of the questionnaire, the effect of question length on number of items reported was modest and not statistically significant. Several reasons for a minimization of the expected effects of question length in this study were proposed earlier.

Since reinterview was another variable to be investigated in the study, 50 percent of the respondents selected for original interview were

designated to be contacted 2 weeks later.* The content of the followup was very close to that of the original interview. Some new questions about health insurance were introduced to avoid excessive repetition. The questionnaire contained the identical chronic conditions checklist. All respondents who originally had been given short-question interviews were given short-question reinterviews; respondents originally asked long questions were reinterviewed with long questions. Thus, data were available to examine potential variations in agreement rates from the first interview to the later one. While agreement rates were poorer the second time than they were originally for both short- and long-question treatments, the deterioration in validity was lower with long questions than it was with short questions.

Since one focus of this study is question length rather than reinterview, tables 46 and 47 show the same type of agreement rates on presence and absence of chronic conditions as presented in tables 44 and 45, but computed this time on the basis of the reinterview data.

*For reasons of field efficiency, all respondents originally interviewed during the first half of the data collection period were designated for reinterview.

Table 46. Probability of physician-respondent agreement on the presence of chronic conditions, by type of agreement rate and question length (reinterview)

Type of agreement rates	Questionnaire procedure[1]		Difference	Percent increase due to question length
	Short questions	Long questions		
$\frac{A}{A+B}$ Probability that chronic conditions checked as present by the physician were reported as present by the respondent	.533	.597	.064	+12
$\frac{A}{A+C}$ Probability that chronic conditions reported as present by the respondent were checked as present by the physician	.408	.527	[2].119	+29
$\frac{A}{A+B+C}$ Overall probability of agreement between physician and respondent for chronic conditions mentioned as present by either of them	.301	.389	[2].088	+29

[1]Number of persons reinterviewed were 49 in short-question procedure and 53 in long-question procedure.
[2]$p < .05$, one-tailed, based on Z.

Table 47. Probability of physician-respondent agreement on the absence of chronic conditions, by type of agreement rate and question length (reinterview)

Type of agreement rates	Questionnaire procedure[1]		Difference	Percent increase due to question length
	Short questions	Long questions		
$\frac{D}{D+C}$ Probability that chronic conditions checked as absent by the physician were reported as absent by the respondent	.869	.894	[2].025	+3
$\frac{D}{D+B}$ Probability that chronic conditions reported as absent by the respondent checked as absent by the physician	.916	.919	.003	([3])
$\frac{D}{D+B+C}$ Overall probability of agreement between physician and respondent for chronic conditions mentioned as absent by either of them	.804	.829	.025	+3

[1]Number of persons reinterviewed were 49 in short-question procedure and 53 in long-question procedure.
[2]$p < .10$, one-tailed, based on Z.
[3]More than zero, but less than 1 percent.

All three agreement rates shown in table 46 were substantially improved by the use of long questions in reinterview. The obtained increase in two of the agreement rates reaches statistical significance ($p \leq .05$). The increase in overall probability of agreement between physician and respondent on the presence of chronic conditions amounts to 29 percent.

While increases in agreement on the absence of conditions due to question length are very modest (table 47), there is indication again that no deterioration occurred for this type of agreement as a result of question length in reinterviews. For the first agreement rate, the increase obtained in long-question reinterviews reached statistical significance at the 10-percent level. Thus, the data obtained the second time confirm the findings obtained originally. Long questions elicit a more valid report than do short questions under conditions of reinterview, as well as under conditions of initial interview.

In summary, this second experiment demonstrated that improvement in the validity of the reporting of health conditions can be achieved without changing the content or meaning of the questions, but only by increasing their length. Somehow the retrieval of more accurate information is facilitated by a question of long duration. The request for accurate reporting is implicitly conveyed by long questions, even though the explicit demand for information and apparently the response duration are unchanged.

CONCLUSIONS

On the basis of the data presented, three major suggestions have been proposed which can be summarized as follows: When the length of survey interview questions is substantially increased and their information demand held constant (a) no appreciable increase is obtained in response duration; yet (b) the response contains more information; and (c) the reported information is more valid.

The first suggestion contradicts other research in which it was found that there was a speech-length matching effect in the interview. This matching effect in other research might have resulted from an uncontrolled increase in the information demanded by long questions. The present study indicates that, when information demand is controlled, long questions do not elicit long responses. Experimental manipulations of information demanded, associated with a control imposed on question length, might help to solve the issue in further research.

That long questions in comparison with short ones might elicit more information and a more accurate report is contradictory to common assumptions and current survey methodology. However, the evidence in this paper leads to the conclusion that lengthy and redundant questions, as designed in this study, elicit increased accuracy of report, even though the responses do not last any longer than those in response to short questions. These findings are puzzling, and they raise questions of importance to survey research. The following should be considered as results of preliminary investigations that require replication:

a. The length of the question has cueing effects upon reporting behavior, causing increased accuracy but not extending to response duration. A long question may convey to the respondent the idea that, because the interviewer has spent much time asking the question, the task is important and, therefore, requires serious efforts. Furthermore, a long question may indicate to the respondent that the interviewer is not in a hurry, and thus releases perception-of-time constraints possibly detrimental to adequate performance in regular interviews. Finally, the responding behavior may also gain in effectiveness because some of the initial ruminating-type activity has already taken place while the question was being asked. A long question may therefore provide cues leading to more adequate performance and at the same time prepare the respondent for the expression of more efficient verbal behavior.

b. Question length or interviewer speech duration is only a vehicle or a proxy for another influential variable, namely, the time given for recall activity. A long question increases the time available for search activity and thus improves the outcome of recall.

c. Since increases in question length have been implemented partially by introducing redundancy in question wording, the multiple presentation of the stimulus may be the influential variable. It may act either because of increased exposure time or because of a repeated

trials effect. Finally, it may be that redundancy improves the clarity of the question which also leads to better reporting performance.

According to this analysis, the effects of at least three variables should be investigated in further research: question length per se and its cue-giving properties; time provided by the question for recall activity; and redundancy of the question. In the experiments described earlier in this paper, all three dimensions have been varied simultaneously so that the specific effect of each could not be isolated. In the experimental treatments, questions were longer; they also provided more time for recall since their first statement always referred to a major part of the question content; and they were redundant.

The following design proposal is an attempt to investigate the specific effects of variations in total question length and recall time, while partially controlling for redundancy. Experimental questions could be designed using the following pool of statements—equal in length—according to various arrangements:

Q = *question* in its short standard form;
F = *filler statement* introducing extraneous information of inconsequential character and unrelated to the specific question demand; and
q = *introductory statement* describing the topic of the question in a manner sufficient to stimulate relevant recall activity.

Each experimental questionnaire would contain only one type of question structure, as described below:

Questionnaire 1 = Q: questions in their short standard form. Question length and recall time are low.

Questionnaire 2 = FQ: the short question preceded by a filler statement. The total length of the question is roughly doubled, whereas the time for recall activity is unchanged since the filler is entirely unrelated to the question demand. Question length is medium and recall time is low.

Questionnaire 3 = qQ: the short question preceded this time by a statement introducing the major question demand. Question length and recall time are both medium.

Questionnaire 4 = FqQ: the short question is preceded by the introductory statement, which is itself preceded by an irrelevant filler, question length is high, and recall time is medium.

Questionnaire 5 = qFQ: the short question is preceded by a filler, which is itself preceded by an introductory statement. Question length and recall time are both high.

Redundancy occurs whenever a question uses both *q* and *Q*, so that this variable is controlled within groups of treatments 1-2 where there is no redundancy and within groups 3-4-5 where redundancy has been introduced.

This experimental design is presented in figure 2, where each cell represents one questionnaire procedure and indicates the strategy used in question wording for this procedure. Comparison of data from cells 1 and 2 will detect the effect of increased length under conditions of equal recall time and no redundancy. Comparison of cells 3 and 4 will detect an effect of increased length with equal recall time and redundancy. A comparison of cells 2 and 3 will explore a combined effect of increased recall time and redundancy, under conditions of equal length. Comparison of cells 4 and 5 will detect

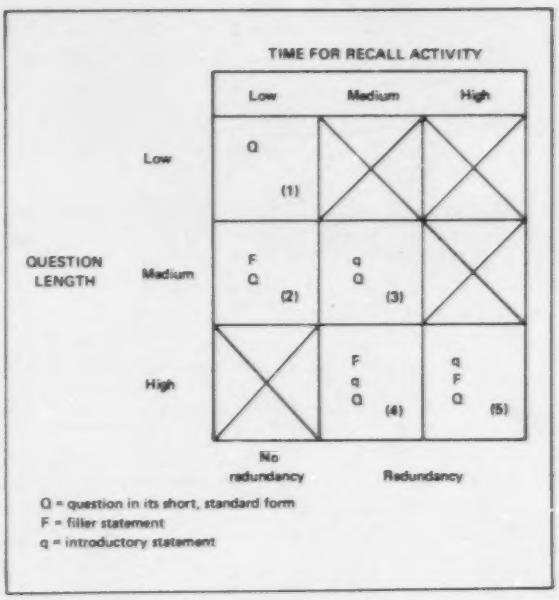


Figure 2. Proposal for a research design.

the effect of increased recall times with equal length and equal redundancy. Diagonal comparisons of cells will provide a means of examining the effects of various combinations of the three variables. The application of this design may generate three types of outcomes:

a. A replication of the earlier findings;

b. An identification of the most efficient strategy or question anatomy; and

c. Some identification of the variables causing improvement in reporting.

REFERENCES

[1]Cannell, C. F., and Fowler, F. J.: *A Study of the Reporting of Visits to Doctors in the National Health Survey.* Ann Arbor, Mich. Survey Research Center. The University of Michigan, Oct. 1963. Mimeographed.

[2]Marquis, K. H., and Cannell, C. F.: *A Study of Interviewer-Respondent Interaction in the Urban Employment Survey.* Ann Arbor, Mich. Survey Research Center. The University of Michigan, Aug. 1969.

[3]National Center for Health Statistics: Interview data on chronic conditions compared with information derived from medical records. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 23. Public Health Service. Washington. U.S. Government Printing Office, May 1967.

[4]National Center for Health Statistics: Health interview responses compared with medical records. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 7. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

[5]Rice, S. A.: Contagious bias in the interview. *Am. J. Sociol.* XXXV: 420-423, 1929.

[6]Stanton, F., and Baker, K. H.: Interview bias and the recall of incompletely learned materials. *Sociometry* 5: 123-124, 1942.

[7]Katz, D.: Do interviewers bias poll results? *Public Opin. Q.* 6: 248-268, 1942.

[8]Robinson, D., and Rohde, S.: Two experiments with an anti-Semitism poll. *J. Abnorm. Soc. Psychol. Q.* 41: 136-144, 1946.

[9]Myers, R. J.: Errors and bias in the reporting of ages in census data. *Trans. Actuarial Soc. Am.* 41: 359-415, 1940.

[10]Neely, T.: *A Study of Error in the Interview.* Privately printed, 1937.

[11]Perry, H., and Crossley, H.: Validity of responses to survey questions. *Public Opin. Q.* 14, 1950.

[12]Lansing, J. B., Ginsberg, G. P., and Braaten, K.: *An Investigation of Response Error.* Urbana, Ill. Bureau of Economic and Business Research, University of Illinois, 1961.

[13]Ferber, R.: *Collecting Financial Data by Consumer Panel Techniques.* Urbana, Ill. Bureau of Economic and Business Research, University of Ilinois, 1959.

[14]Lansing, J. B., and Blood, D. M.: *The Changing Travel Market.* Ann Arbor, Mich. Survey Research Center, The University of Michigan. Monograph No. 38, 1964.

[15]Goddard, K. E., Broder, G., and Wenar, C.: Reliability of pediatric histories, a preliminary study. *Pediatrics* 28 1011-1018, 1961.

[16]Weiss, D. J., et al.: Validity of work histories obtained by interview. *Minn. Stud. Vocat. Rehabil.* 12(41): 1961.

[17]National Center for Health Statistics: Reporting of hospitalization in the Health Interview Survey. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 6. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

[18]National Center for Health Statistics: Comparison of hospitalization reporting in three survey procedures. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 8. Public Health Service. Washington. U.S. Government Printing Office, July 1965.

[19]Thorndike, E. L.: *Educational Psychology.* New York. Teachers College Press, Columbia University, 1913.

[20]Ebbinghaus, H.: *Memory.* Translated by H. A. Ruger and C. E. Bussenius. New York. Teachers College, 1913. (Reissued as paperback, New York: Dover, 1964.)

[21]Koffka, K.: *Principles of Gestalt Psychology.* New York. Harcourt Brace, 1935.

[22]Köhler, W.: *Gestalt Psychology.* New York. Liveright, 1947.

[23]McGeoch, J. A.: Forgetting and the law of disuse. *Psychol. Rev.* 39: 352-370, 1932.

[24]Postman, L.: The present status of interference theory, in C. N. Cofer, ed., *Verbal Learning and Verbal Behavior.* New York. McGraw-Hill, 1961. pp. 152-179.

[25]Hilgard, E. R., and Bower, G. H.: *Theories of Learning.* 3rd ed. New York. Appleton, 1966.

[26]Cobb, S., and Cannell, C. F.: Some thoughts about interview data. *Int. Epidemiol. Assoc. Bull.* 13: 43-54, 1966.

[27]Taffel, C.: Anxiety and the conditioning of verbal behavior. *J Abnorm. Soc. Psychol.* 51: 496-501, 1955.

[28]Greenspoon, J.: The reinforcing effect of two spoken sounds on the frequency of two responses. *Am. J. Psychol.* 68: 409-416, 1955.

[29]Verplanck, W. S.: The control of the content of conversation: Reinforcement of statements of opinion. *J. Abnorm. Soc. Psychol.* 51: 668-676, 1955.

[30]Azarin, N. H., et al.: The control of conversation through reinforcement. *J. Exp. Anal. Behav.* 4: 25-30, 1961.

[31]Centers, R.: A laboratory adaptation of the conversational procedure for the conditioning of verbal operants. *J. Abnorm. Soc. Psychol.* 67: 334-339, 1964.

[32]Kanfer, F. H., and McBrearty, J. F.: Minimal social reinforcement and interview content. *J. Clin. Psychol.* 18: 210-215, Apr. 1962.

[33]Hildrum, D. C., and Brown, R. W.: Verbal reinforcement and interviewer bias. *J. Abnorm. Soc. Psychol.* 53: 108-111, 1956.

[34]Nuthman, A. M.: Conditioning of a response class on a personality test. *J. Abnorm. Soc. Psychol.* 54: 19-23, 1957.

[35]Staats, A. W., et al.: Operant conditioning of factor analytic personality traits. *J. Gen. Psychol.* 66: 101-114, Jan. 1962.

[36]Singer, R. D.: Verbal conditioning and generalization of prodemocratic responses. *J. Abnorm. Soc. Psychol.* 63: 43-46, July 1961.

[37]Insko, C. A.: Verbal reinforcement and attitude. *J. Pers. Soc. Psychol.* 2: 621-623, Oct. 1965.

[38]Kahn, R. L., and Cannell, C. F.: *The Dynamics of Interviewing.* New York. John Wiley & Sons, Inc., 1957.

[39]Cannell, C. F., and Kahn, R. L.: Interviewing, in G. L. Lindzey and E. Aronson, eds., *The Handbook of Social Psychology.* Reading, Mass. Addison Wesley, 1968.

[40]National Center for Health Statistics: International comparisons of medical care utilization: A feasibility study. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 33. Public Health Service. Washington. U.S. Government Printing Office, June 1969.

[41]Mooney, H. W.: *Methodology in Two California Health Surveys.* Public Health Monograph No. 70. PHS Pub. No. 942. Public Health Service. Washington. U.S. Government Printing Office, 1962.

[42]National Center for Health Statistics: Optimum recall period for reporting persons injured in motor vehicle accidents. *Vital and Health Statistics.* Series 2-No. 50. DHEW Pub. No. (HSM) 72-1050. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, Apr. 1972.

[43]World Health Organization: *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death,* Based on the Recommendations of the Seventh Revision Conference, 1955. Geneva. World Health Organization, 1957.

[44]Fowler, F. J., Jr.: Education, Interaction, and Interview Performance. Doctoral dissertation, The University of Michigan, 1965.

[45]Weiss, C. H.: *Validity of Interview Responses of Welfare Mothers.* New York. Bureau of Applied Social Research, Columbia University, Feb. 1968.

[46]National Center for Health Statistics: The influence of interviewer and respondent psychological and behavioral variables on the reporting in household interviews. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 26. Public Health Service. Washington. U.S. Government Printing Office, March 1968.

[47]Mueller, J. H., Schuessler, K. F., and Costner, H. L.: *Statistical Reasoning in Sociology.* Boston. Houghton-Mifflin, 1970. pp. 279-291.

[48]deKoning, T. L.: Interviewer and Respondent Interaction in the Household Interview. Doctoral dissertation, The University of Michigan, 1969.

[49]Marshall, J., Marquis, K. H., and Oskamp, S.: Effects of kind of question and atmosphere of interrogation on accuracy and completeness of testimony. *Harv. Law Rev.* 84(7): 1620-1643, May 1971.

[50]National Center for Health Statistics: Effect of some experimental interviewing techniques on reporting in the Health Interview Survey. *Vital and Health Statistics.* PHS Pub. No. 1000-Series 2-No. 41. Public Health Service. Washington, U.S. Government Printing Office, May 1971.

[51]Dollard, J., and Miller, N. E.: *Personality and Psychotherapy.* New York. McGraw-Hill, 1950.

[52]Skinner, B. F.: *Verbal Behavior.* New York. Appleton Century Crofts, 1957.

[53]National Center for Health Statistics: Reporting of health events in household interviews: Effects of reinforcement, question length, and reinterviews. *Vital and Health Statistics.* Series 2-No. 45. DHEW Pub. No. (HSM) 72-1028. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, Mar. 1972.

[54]Marquis, K. H., and Laurent, A.: Accuracy of Health Reporting in Household Interviews as a Function of Reinforcement, Question Length, and Reinterviews. Unpublished manuscript, 1969.

[55]Cannell, C. F., Fowler, F. J., and Marquis, K. H.: *Respondents Talk About the National Health Survey Interview.* Ann Arbor, Mich. Survey Research Center. The University of Michigan, Mar. 1965. Mimeographed.

[56]Cannell, C. F., Fowler, F. J., and Marquis, K. H.: Report on Development of Brochures for H.I.S. Respondents. Ann Arbor, Mich. Survey Research Center. The University of Michigan, Mar. 1965. Unpublished manuscript.

[57]Zajonc, R. B.: *Social Psychology, An Experimental Approach.* Belmont, Calif. Wadsworth, 1966. (*Basic Concepts in Psychology* series.)

[58]Edwards, A. L. *The Social Desirability Variable in*

Personality Assessment and Research. New York, Dryden Press, 1957.

[59]Bales, R. F.: *Interaction Process Analysis.* Cambridge, Mass. Wesley Pre , 1950.

[60]Hyman, H. H., t al.: *Interviewing in Social Research.* Chicago. University of Chicago Press, 1954.

[61]Cannell, C. F.: Private communication: Proposal. 1969.

[62]Krasner, L.: Studies of the conditioning of verbal behavior. *Psychol. Bull.* 55: 3, 1958.

[63]Spielberger, C. D.: The role of awareness in verbal conditioning, in C. W. Erikson, ed. *Behaviorism and Awareness.* Durham, N.C. Duke University Press, 1962, pp. 73-101.

[64]Spielberger, C. D., and DeNike, L. D.: Descriptive behaviorism vs. cognitive theory in verbal operant conditioning. *Psych. Rev.* 73(2): 306-326, 1966.

[65]Dulany, D. E.: Hypotheses and habits in verbal "operant conditioning." *J. Abnorm. Soc. Psychol.* 63: 251, 263, 1961.

[66]Oakes, W. F.: Verbal operant conditioning, intertrial activity, awareness and the extended interview. *J. Pers. Soc. Psychol.* 6: 198-202, 1967.

[67]Marquis, K. H.: *Report of Legal Study.* Ann Arbor, Mich. Survey Research Center, The University of Michigan, 1967. Unpublished manuscript.

[68]Asch, S. E.: A reformulation of the problem of associations. *Am. Psychol.* 24: 92-102, 1969.

[69]National Center for Health Statistics: Reporting health events in household interviews: Effects of an extensive questionnaire and a diary procedure. *Vital and Health Statistics.* Series 2-No. 49. DHEW Pub. No. (HSM) 72-1049. Health Services and Mental Health Administration. Washington. U.S. Government Printing Office, Apr. 1972.

[70]Parten, M.: *Surveys, Polls, and Samples: Practical Procedures.* New York. Harper & Bros., 1950.

[71]Oppenheim, A. N.: *Questionnaire Design and Attitude Measurement.* New York. Basic Books, 1966.

[72]Matarazzo, J. D., Saslow, G., and Wiens, A. N.: Studies in interview speech behavior, in L. Krasner and L. P. Ullman, eds., *Research in Behavior Modification.* New York. Holt, Rinehart and Winston, 1965. pp. 179-210.

[73]Matarazzo, J. D., Wiens, A. N., Saslow, G., Dunham, R. H., and Voas, R. B.: Speech durations of astronaut and ground communicator. *Science* 143: 148-150, 1964.

[74]Ray, M. L., and Webb, E. J.: Speech duration effects in the Kennedy news conferences. *Science* 153: 899-901, 1966.

[75]Heller, K., Davis, J. D., and Myers, R. A.: The effects of interviewer style in a standardized interview. *J. Consult. Psychol.* 30(6): 501-508, 1966.

APPENDIX

APPLICATION OF SURVEY RESEARCH CENTER FINDINGS TO HEALTH INTERVIEW SURVEY PROCEDURES

Based on the experience of many researchers in the health field and other areas it was expected that underreporting would be one of the major problems to be solved in the data collection phase of the Health Interview Survey. However, before any attempt could be made to remedy this shortcoming of the interview method, it was necessary to obtain some idea of the magnitude of the problem, to learn something about the characteristics of persons who fail to report health events, and to identify the particular kinds of events that tend to be underreported. The need for this kind of information led to a number of studies involving a comparison of information provided by interview respondents with independent medical records of known validity. Several of these studies, conducted by the Survey Research Center (SRC) of the University of Michigan, provided information that led to major revisions in the collection and processing of Health Interview Survey (HIS) data.

Recall of Health Events

One of the early studies carried out by SRC in 1958-59 consisted of a comparison of hospitalizations reported in interviews with actual hospital records. A sample of patients discharged from hospitals participating in the Professional Activity Study were interviewed, and the results were compared with the discharge records (see ref. 17 for a complete description of this study). Findings of this study, described earlier in this report, demonstrate that underreporting of hospitalization in a health interview situation is influenced by the impact of the hospitalization, the threat or embarrassment caused by the nature of the condition causing hospitalization, and the time elapsed between the interview and the period spent in the hospital. Since the first two of these findings involved intrinsic characteristics of the health event, no immediate solution to underreporting associated with these causes was available.

However, the consistent increase in underreporting with the time elapsed between hospitalization and the interview was a finding that appeared to have practical application to data collected in the Health Interview Survey. As a result, in the derivation of estimates of the volume of hospital discharges from the basic HIS data collected beginning in July 1958, the 6-month period preceding the date of interview was used as the period of reference. By doubling the weight attached to each of the reported events within that period, it was possible to produce estimates comparable to those based on 12 months of recall, but with considerably less of the underreporting bias introduced by the use of the longer recall period.

The use of the 12-month recall period was continued in the collection of data on hospital experience because of the several kinds of estimates produced from the Survey. To combine the hospital episodes of sample persons in order to estimate the number of persons with one or more episodes in a given year, it is necessary to consider a year's experience for each sample person. On the other hand, in estimating the annual volume of hospital discharges, any recall period can be used if the weight attached to each event in the estimating

procedure is properly adjusted. Since the length of the recall period is inversely related to the magnitude of the sampling error, the 6-month reference period was selected so that response bias could be appreciably reduced without an undue increase in the size of the sampling error.

The imprecision with which respondents recalled dates of health events during an interview, brought to light by several of the SRC studies, led to the use of a recall period in the collection phase of hospital data extending beyond the year preceding the interview. For example, persons interviewed during July of a particular year were asked about their hospital experience since May of the previous year. This innovation improved the reporting of events that occurred near the beginning of the reference year, as well as hospitalizations that started prior to the year preceding the interview but extending into the reference year. During the processing phase, those hospitalizations for which no days during the year prior to interview were recorded were eliminated from the hospitalization data.

The SRC record-check studies also revealed inaccuracies in the reporting of physician visits. Even though the recall period for the reporting of physician visits was limited to the 2-week period prior to week of interview, some of the visits occurring during that period were not reported and, in other instances, visits occurring prior to the period were reported as happening within the recall period. This finding eventually led to the decision to enumerate physician visits on the questionnaire by date of visit so that comparison of the occurrence and interview dates would establish that the event had occurred during the appropriate recall period.

Effective Probing for Health Events

Several of the Survey Research Center studies have indicated that much of the information not reported during an interview has not been repressed, nor has it disappeared from memory. It was simply not elicited because the questioning procedures failed to bring it forth. This finding suggested that the probe questions designed to encourage the reporting of health events in the HIS were not stimulating retrieval of information sufficiently, and that questions constructed to elicit certain kinds of response should be added.

Prior to July 1962, respondents in the HIS were asked about overnight stays in hospitals of family members during the previous 12 months, and about stays in nursing homes and sanitariums. Comparison of the estimates of hospitalizations for delivery derived from these data with natality figures for the years 1958-62 indicated that hospitalizations of this type were underreported in the interview survey. To correct this situation, a probe question directed particularly to the population at risk was added to the questionnaire. In households where children 1 year of age or under were reported as household members, the following probe questions were asked: "When was (the child) born?" "Was (the child) born in a hospital?" "Is this hospitalization included in the number you gave me?" If the hospitalization had occurred during the reference period and it had not been reported in response to earlier probes, then the entries on the questionnaire for the mother and the child were corrected. The addition of this series of questions resulted in an appreciable decrease in the amount of underreporting in this area of the questionnaire.

Through June 1964, the reporting of information on the number of physician visits during the 2 weeks prior to week of interview was dependent on one probe question: "Last week or the week before, did anyone in the family talk to a doctor or go to a doctor's office or clinic?" Beginning in January 1966, the next period during which information on physician visits was collected in the Survey, two probe questions were added: "During that 2-week period has anyone in the family been to a doctor's office or clinic for shots, X-rays, tests, or examinations?" and, "During that period, did anyone in the family get any medical advice from a doctor over the telephone?" The first of these questions was added to remind the respondent of visits that were made for preventive care or, in some instances, for reasons other than treatment of illness. The second question informed the respondent that telephone calls to obtain medical advice were considered as physician vists in the Survey. Both of these questions had been used by SRC in the study designed to evaluate interviewer performance over time.

Because these probe questions were added at the same time as the question regarding the date

of the visit, their effect on the data was not as obvious as it otherwise would have been. The procedure of relating the date of the occurrence of a visit to the date of interview to determine if it actually occurred during the proper recall period effectively excluded all overreporting of visits that had actually occurred prior to the recall period or during the interview week. Previously such visits would have compensated for some of the underreporting. Their removal from the data made it difficult to evaluate the effectiveness of the added probe questions in terms of additional visits reported, but there is evidence that the yield from these questions was substantial.

Interviewer-Respondent Communication

Many of the studies conducted by SRC have emphasized the importance of the influence the interviewer exerts on the respondent and, in turn, on the completeness and accuracy of the reported data. The interviewer's attitude, her expectations, the kind of feedback she provides, and her behavior during the interview are only a few of the factors that determine the kind and amount of information obtained during an interview. However, to take advantage of this phenomenon in order to improve the quality of reported information, controls must be initiated to avoid the introduction of interviewer biases. One of the best methods of exercising control of interviewer behavior is to include devices, questions, and statements in the questionnaire which will improve communication between the participants in the interview but will not direct the responses.

Some of the innovations in the HIS questionnaire that have resulted from this type of research include the following:

a. A simple introductory statement has been prepared in which the interviewer identifies herself at the door of the household and explains very briefly the purpose of her visit. In case the respondent (or another family member) wants to know more about the purpose of the survey or the uses of the collected data, a more detailed statement is available to the interviewer.
b. Within the questionnaire, introductory statements are used to explain the subject matter about which questions are to be asked and to serve as transition devices from one health topic to another. For example, a section on X-ray visits was introduced by the statement, "Exposure to all kinds of X-rays is a matter of particular interest to the Public Health Service, and I have some questions about X-rays and fluoroscopes."
c. A small calendar card, with the appropriate 2-week recall period outlined in red is handed to the respondent early in the interview so that she is constantly aware of the specific 2-week period referred to throughout the interview.
d. Nondirective probe questions have been included on the questionnaires in areas where nonspecific or ambiguous information is likely to be reported. For example, if the respondent reports that she visited the doctor at a clinic, the interviewer is instructed to ask: "Was it a hospital outpatient clinic, a company clinic, or some other kind of clinic?"

In the SRC study on interviewer performance over time, it was found that interviewers became less careful or conscientious in using the techniques they were trained to use. There was also evidence that interviewers who performed well inspired their respondents to perform well, as measured by the reporting of hospitalizations. This study also brought to light the need for interviewer training to include devices for stimulating the interviewer's enthusiasm for her job in addition to retraining in the use of interviewing techniques. These findings have been reinforced by records of interviewer performance maintained by the Bureau of the Census, and have been taken into account in the preparation of material for the periodic training and retraining sessions conducted for interviewers in the Health Interview Survey.

Other Considerations

As in any series of research studies, some of the experimental measures in the SRC series, when tested as methods for reducing underreporting in interviews, did not contribute any significant findings. In other instances, encouraging results from studies were either inconclusive

or needed further testing in a nonlaboratory situation. In this latter category is the finding that long questions are more effective than short ones in bringing forth complete and accurate responses. More research is needed to determine if long questions are productive because they have cuing effects on reporting behavior, allow more time for recall activity, or merely because they introduce redundancy of stimuli. Until the specific variables causing improved reporting are identified, the introduction of longer questions on the HIS questionnaire, which would lead to longer interviews and increased costs, could not be justified.

Verbal reinforcement by the interviewer has been shown to have cognitive and motivational effects on the respondent by instilling awareness of respondent task requirements and encouraging adequate responses to subsequent questions. However, it will be necessary to develop ways in which the interviewer can effectively use reinforcement without introducing an undue amount of bias in the collected data before this device can be seriously considered.

All of the studies in which the SRC has attempted to measure the amounts and types of underreporting in interviews indicate that some basic principles of memory and retrieval can be used to improve reporting. Further research is needed on the ways in which information is stored and on effective methods of retrieving that information.

———— ○○○ ————

• U.S. GOVERNMENT PRINTING OFFICE: 1977—241-180:3

VITAL AND HEALTH STATISTICS PUBLICATIONS SERIES

Formerly Public Health Service Publication No. 1000

Series 1. Programs and Collection Procedures.—Reports which describe the general programs of the National Center for Health Statistics and its offices and divisions, data collection methods used, definitions, and other material necessary for understanding the data.

Series 2. Data Evaluation and Methods Research.—Studies of new statistical methodology including experimental tests of new survey methods, studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, contributions to statistical theory.

Series 3. Analytical Studies.—Reports presenting analytical or interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

Series 4. Documents and Committee Reports.—Final reports of major committees concerned with vital and health statistics, and documents such as recommended model vital registration laws and revised birth and death certificates.

Series 10. Data from the Health Interview Survey.—Statistics on illness; accidental injuries; disability; use of hospital, medical, dental, and other services; and other health-related topics, based on data collected in a continuing national household interview survey.

Series 11. Data from the Health Examination Survey.—Data from direct examination, testing, and measurement of national samples of the civilian, noninstitutionalized population provide the basis for two types of reports: (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics; and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

Series 12. Data from the Institutionalized Population Surveys.—Discontinued effective 1975. Future reports from these surveys will be in Series 13.

Series 13. Data on Health Resources Utilization.—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

Series 14. Data on Health Resources: Manpower and Facilities.—Statistics on the numbers, geographic distribution, and characteristics of health resources including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

Series 20. Data on Mortality.—Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death, age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records, based on sample surveys of those records.

Series 21. Data on Natality, Marriage, and Divorce.—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records, based on sample surveys of those records.

Series 22. Data from the National Mortality and Natality Surveys.—Discontinued effective 1975. Future reports from these sample surveys based on vital records will be included in Series 20 and 21, respectively.

Series 23. Data from the National Survey of Family Growth.—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a biennial survey of a nationwide probability sample of ever-married women 15-44 years of age.

For a list of titles of reports published in these series, write to: Scientific and Technical Information Branch
National Center for Health Statistics
Public Health Service, HRA
Rockville, Md. 20857